DOCUMENT RESUME

ED 041 263 24 AL 002 464

AUTHOR	Willis, Clodius
TITLE	The Development of an Automatic Dialect Classification Test. Final Report.
INSTITUTION	Rochester Univ., N.Y.
SPONS AGENCY	Institute of International Studies (DHEW/OE), Washington, D.C.
BUREAU NO	BR-8-B-108
PUB DATE	Jul 69
GRANT	OEG-2-942-10801006(010)
NOTE	181p.
EDRS PRICE	EDRS Price MF-$0.75 HC-$9.15
DESCRIPTORS	Acoustic Phonetics, American English, *Artificial Speech, *Auditory Perception, *Dialect Studies, English, Perception, Phonology, *Regional Dialects, *Vowels

ABSTRACT

These experiments investigated and described intra-subject, inter-subject, and inter-group variation in perception of synthetic vowels as well as the possibility that inter-group differences reflect dialect differences. Two tests were made covering the full phonetic range of English vowels. In two other tests subjects chose between one of two vowels, specifically the vowels in the contrasts bet/bat, cot/caught, and hat/hot. The tests were administered to 196 subjects from four dialect regions: Western New York, Virginia, Ontario (Canada), and Massachusetts. The data which resulted from the tests gave rise to the following conclusions: (1) There is ample evidence for the existence of dialectally determined variation in vowel preception. (2) Specific inter-group differences which seem clearly to correlate with dialect are the particular treatment of /a/ before /r/ by Massachusetts subjects, differences in perception of the contrast bet/bat, and differences in position of the phoneme boundaries /i/-/e/, /ae/-/a/, and /o/-/u/. (3) Certain results in the data, such as the contrast of cot/caught, cannot easily be explained by dialect differences. The report concludes with general observations on performance and reliability of tests of this type, possible applications of the tests to dialect study, and theoretical insights afforded by the study. (Author/JD)

BR 8-B-108
PA 24

ED041263

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT
Project No. 8-B-108
Grant No. OEG-2-9-420108-1006-(010)

THE DEVELOPMENT OF AN AUTOMATIC DIALECT CLASSIFICATION TEST

Clodius Willis
The University of Rochester
River Campus Station
Rochester, New York 14627

July, 1969

AL 002 464

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Project No. 8-B-108
Grant No. OEG-2-9-420108-1006-(010)

THE DEVELOPMENT OF AN AUTOMATIC DIALECT CLASSIFICATION TEST

Clodius Willis

The University of Rochester

Rochester, New York 14627

July, 1969

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

VITAE

Date of Birth: August 4, 1937

Place of Birth: Schenectady, New York

Other Degrees Awarded:

B.A., Amherst College, 1959, a major in French Language and Literature

M.A.T., Harvard University, 1963, a major in French Language and Literature

Academic Connections:

The University of Rochester
Residence, 1964-1967
Teaching Assistant, 1964-1967
Advisor, Dr. Dean H. Obrecht

Canisius College
Instructor in French, 1967-1968

Clark University
Assistant Professor of Romance Languages, 1968-1969

ACKNOWLEDGEMENTS

I wish to express my gratitude to Dean Obrecht. Without his friendship and advice, this thesis would not have been possible.

This thesis is dedicated in recognition of the debt I owe to my parents, Clodius Harris Willis and Katherine Vaughan Willis, and in their memory.

The work described here has been supported in part by the Office of Education in connection with Contract OEG-2-9-420108-1006-(010).

ABSTRACT

The purpose of these experiments is to investigate and describe intra-subject, inter-subject, and inter-group variation in perception of synthetic vowels. The possibility that inter-group differences reflect dialect differences is also investigated.

Part I, an introductory chapter begins with a review of previous experiments in perception of synthetic vowels, which have largely been concerned with the average response of speakers of a given language, and with the discovery of the mechanism of speech perception. The two formant description of vowels, the psychological reality of phoneme centers, and the relation between perception and articulation of language are discussed in pointing out the limitations of synthetic vowel perception tests. There is also a discussion of cross-linguistic studies, and previous studies of individual and dialectally determined variation in vowel perception.

Part II describes the methods of investigation, beginning with a precise description of synthesized vowels. There are two tests covering the full phonetic range of English vowels, and four tests in which the subjects choose between one of two vowels, specifically the vowels in the contrasts bet-bat, cot/caught, and hat/hot. These tests

were administered to 196 subjects from four dialect regions; Western New York, Virginia, Ontario, Canada, and Massachusetts. The data were tabulated by computer and are displayed in the Appendices, in the form of percentages of response, for each category and each stimulus.

Part III is an analysis of the data. Individual and group performance, as well as group variability over time, are given precise definitions, and measured for each test, and for all groups. There is finally an analysis of inter-group differences, and of their reliability and statistical significance. From these analyses, the following conclusions are drawn: 1) There is ample evidence for the existence of dialectally determined variation in vowel perception; 2) Specific inter-group differences which seem clearly to correlate with dialect are the particular treatment of /a/ before /r/ by Massachusetts subjects, differences in perception of the contrast bet/bat, and differences in position of the phoneme boundaries /i/-/e/,/ae/-/a/, and /o/-/u/. 3) Certain results in the data cannot easily be explained by dialect differences. Results of the test contrasting cot/caught are a particular problem.

Part IV contains more general conclusions, pertaining to performance and reliability with tests of this type, to possible applications of the tests to dialect study, and to theoretical insights afforded by the study.

TABLE OF CONTENTS

I. INTRODUCTION 1

A. A DESCRIPTION OF SYNTHETIC VOWEL PERCEPTION TESTS 1

B. IMPORTANT ASSUMPTIONS OF SYNTHETIC VOWEL PERCEPTION TESTS 2

C. THE RELATION BETWEEN PERCEPTION OF SYNTHETIC VOWELS AND PERCEPTION OF NATURAL LANGUAGE 11

II. METHOD OF INVESTIGATION 21

A. AN OVER-VIEW OF THE EXPERIMENTS 21

B. PRODUCTION OF THE SYNTHETIC VOWEL STIMULI .. 23

C. THE COMPOSITION OF THE TESTS 26

D. ADMINISTRATION OF THE TESTS 28

E. DESCRIPTION OF THE SUBJECTS 29

F. DESCRIPTION OF THE DATA 32

III. RESULTS 36

A. MAJOR RESPONSE CATEGORIES 36

B. COMMON PHONEME CENTERS 36

C. PERFORMANCE 40

D. VARIABILITY 48

E. INTER-GROUP DIFFERENCES 59

1. Concerning Methods of Analysis 59

2. Differences in Selection of Phoneme Centers (Test IB) 61

3. Inability to Categorize Certain Sets of Stimuli (Tests IIA, IIB) 63

4. Differences in Position of Phoneme Boundaries 65

F. EXPLANATION OF INTER-GROUP DIFFERENCES BY REFERENCE TO DIALECTAL FEATURES 85

1. Previously Established Features of the Dialects of Western New York, Massachusetts, Ontario, and Virginia 85

2. A Summary of Inter-group Differences 87

3. Correlation of Inter-group Differences and Dialect Features 89

IV. CONCLUSIONS 95

A. PERFORMANCE 95

B. VARIABILITY 96

C. PRACTICAL APPLICATIONS OF SYNTHETIC VOWEL PERCEPTION TESTS TO DIALECT STUDY 98

D. THEORETICAL VALUE OF THE STUDY OF VARIATION IN PERCEPTION OF SYNTHETIC VOWELS 99

BIBLIOGRAPHY 102

APPENDICES 107

Appendix A - The Production of the Stimuli 107

Appendix B - The Composition of the Tests 108

Appendix C - The Administration of the Tests 115

Appendix D - The Subjects 118

Appendix E - Sample Data 127

Appendix F - Results of Test IA 130

Appendix G - Results of Test IB 135

Appendix H - Results of Test IIA 145

LIST OF TABLES IN THE TEXT

TABLE I - Explanation of Notation Used in Computer Output 34

TABLE II - Major Response Categories 37

TABLE III - Performance on Test IB 43

TABLE IV - Response Percentages for Phoneme Centers for Test IB 45

TABLE V - Percentages of Good Patterns 47

TABLE VI - Performance and Sex 49

TABLE VII - Performance and I.Q. 49

TABLE VIII - Variability for Test IIIC 57

TABLE IX - Differences in Selection of Phoneme Centers 62

TABLE X - Responses to Stimulus Number 29, Test IB , 66
Number of /i/ and /I/ Responses 66

TABLE XI - Responses to Stimulus Number 29. Test IB, Number of /e/ Responses 67

TABLE XII - Responses to Stimulus Number 18, Test IB 71

TABLE XIII - Percentages of /ae/ Response, for Six Stimuli from Test IIC 77

TABLE XIV - Number of /ae/ Responses to Stimulus 12, Test IIC 78

TABLE XV - Percentages of /a/ Response, for Six Stimuli from Test IIB 82

TABLE XVI - Number of /a/ Responses to Stimulus 5 and Stimulus 29, Test IIB 83

TABLE XVII - Responses to Stimulus 28, Test IB 84

LIST OF FIGURES IN THE TEXT

FIGURE A - Sample of Test Results 33

FIGURE B - Common Phoneme Centers 39

FIGURE C - Correlation of Phoneme Centers with F1/F2 areas for Tests IB, IIA, B, C ... 41

FIGURE D - Variability for Test IB 51

FIGURE E - Variability for Test IIA 52

FIGURE F - Variability for Test IIB 54

FIGURE G - Variability for Test IIC 56

FIGURE H - The Position of the /E/-/ae/ Phoneme Boundary for Test IB 70

FIGURE I - The Position of the /E/-/ae/ Phoneme Boundary for Test IIA 74,75

FIGURE J - The Position of the /ae/-/a/ Phoneme Boundary for Test IIC 76

FIGURE K - The Position of the /ae/-/a/ Phoneme Boundary for Test IIIC 80

FIGURE L - The Position of the /a/-/o/ Phoneme Boundary for Test IIB 81

I. INTRODUCTION

The purpose of this study is to measure the degree of variation in the perception of synthetic vowels. We will describe intra-subject variation and inter-subject variation, as well as inter-group variation. The ultimate usefulness of such measures lies in the possibility that variation is to some extent dialectally determined, and to the extent that this is true, perception of such stimuli may be used as a measure of dialect.

The experiments providing the data for this study are a continuation of a type of experiment in synthetic speech perception largely conceived and implemented by Haskins Laboratories (Cooper, 1961; Liberman, 1957). Since it is this experimental design that we are to use, we shall first examine its assumptions and limitations.

A. A DESCRIPTION OF SYNTHETIC VOWEL PERCEPTION TESTS

Post-war acoustic phonetics established the close relation between the shape of the vocal tract and formant frequencies (Joos, 1948; Fant, 1960). It was also found that vowels could be successfully synthesized with two formants. This discovery led to experiments which investigated the effect of shifts in frequency of the first two formants (F1, F2) on the perception of vowel color. From such experiments came the provocative suggestion that the quality of vowels is specified primarily by formant one and formant two (Delattre, Liberman and Cooper, 1951; Delattre, Liberman, Cooper and Gerstman, 1952).

The stimuli for the Haskins tests under consideration had two formants, and were invariant in every respect but the peak frequencies of these formants. Pitch (F0) and amplitude (A0) were at steady state in all stimuli. From a given area of the F1/F2 chart (the F1/F2 chart being a graph with vertical axis representing F1 frequencies and horizontal axis representing F2 frequencies), a number of stimuli were selected so that they represented a gradation of vowel quality from vowel A to vowel B, adjacent on the traditional vowel triangle. Subjects categorized a random series of these and from the data phoneme boundaries and phoneme centers were determined. We define here phoneme boundaries as the cross-over points at which, as formants shift in acoustical value, vowel A becomes vowel B. A phoneme center, as we shall use the term, is defined as a point of maximum agreement in response category among subjects.

B. IMPORTANT ASSUMPTIONS OF SYNTHETIC VOWEL PERCEPTION TESTS

The synthetic vowel perception tests assume the validity of a two formant description of vowels. By this is meant that while certain insufficiencies of the description are to be admitted, for practical purposes of investigation and description, F1 and F2 are considered as necessary and sufficient cues in the perception of vowels. A closely related implication of a two formant description is that the F1/F2 chart is largely equivalent to the traditional vowel triangle, with its two parameters of tongue height and tongue backing (Joos, 1948, and frequently elsewhere in the literature).

A second important assumption results from the ultimate aim of the Haskins experiments, which was an understanding of a psychological mechanism involved in speech perception. The assumption resulting from this aim is that the phoneme centers and phoneme boundaries as determined in the manner described above are related in a significant, though imperfectly understood way, to some perceptual categories of the subjects taking the tests.

We shall discuss the above assumptions, but since the aim of this study is not the discovery of the mechanism of speech perception, we do not depend directly on the truth of these assumptions. This study is limited to a description of the behavior of subjects confronted with the task of categorizing synthetic vowels, and the correlation of this behavior with other subject variables. In an indirect manner, however, the results of our tests will reflect the nature of speech perception, and give some further evidence for or against such assumptions.

Concerning A Two Formant Description of Vowels

The following discussion of vowel perception is based on authoritative reviews of current speculation on the subject of vowel perception (Flanagan, 1965; Kozhevnikiv and Chistovich, 1966; Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967; Fant, 1967). While there is considerable disagreement as to the nature and size of the basic units of speech perception, all of the above cited authors agree that phonetic features play an important role in the perception of vowels, and we also assume this fact from the outset of this discussion.

There is, however, little agreement as to the precise nature of phonetic cues for vowels. It is not even entirely clear that formant specifications are the correct approach. For a review of arguments for and against formants as the terms of vowel description, see Castle (1964). Experiments in frequency transposition of formants without change in voice fundamental showed that such shifts in formant frequency had little effect on vowel quality, for shifts of as much as 1/2 octave (Oeken, 1963). It is argued that formant peaks are not necessary for vowel-like quality, that, possibly, any complex sound could be identified with a vowel under proper conditions, and that formants cannot always be found in natural vowels of recognizable quality (Carpenter and Morton, 1962). The same authors conclude, however, that while formant peaks are not necessary, they appear to be an important factor when present. This is a sort of compromise view which seems most reasonable to us.

Even if we accept the importance of formants in determining perceived quality of vowels, there are other factors which affect this perception. Allophonic variation and phoneme overlap, as well as contextual, stylistic, and other variations make the one-to-one mapping of phonemes to phonetic features impossible. Formant amplitude and voice frequency play a role in determining vowel color (Peterson, 1961). Hence it appears that formant peak frequencies are perfectly related to vowel color only if other variables are held constant.

Even if it were accepted that formant frequencies are related to vowel color, this is strictly speaking no proof that these formants are the important cues in the perception of natural speech. It has been shown that steady-state formant information is less important for words in context than for isolated vowel sounds (Castle, 1964).

We may ask the further question of how many formants are important. Most work at Haskins Labs considers

F1 and F2 as sufficient for vowel specification. Ladefoged suggests that two formants are sufficient for /e/,/E/,/a/,/ɔ/. (Cited in Morton and Carpenter, 1962, p. 190). Fant, states that F3 is of considerable importance in front vowels. "F2 and higher formants of front vowels appear to constitute a single perceptual cue which plays a role similar to that of an F2 alone in mid and back vowels." (Fant, 1967, p. 11). Only as a cue to centralization of high vowels is F3 a separate cue (Fant and Risberg, 1964).

It is only if we accept a two formant description of vowels that it becomes possible to describe vowels on an F1/F2 graph. While such a graph is a convenient simplification, it is nevertheless a fact that F1 is closely related to degree of tongue height and mouth opening and F2 to the degree of fronting (Fant, 1962). In synthetic vowels, raising F2 has the effect of giving a fronted quality, and raising F1 has the effect of giving a lowered quality to the vowel.

Having pointed out the limitations of a two formant theory of vowels, we conclude by saying that for purposes of a simple description of acoustic, physiological, and perceptual qualities of phonetic features of vowels, an F1/F2 chart seems adequate. An F1/F2 chart bears close relation to the traditional coordinates of fronting, backing, rounding, unrounding, raised, and lowered, and this particularly when dealing with synthetic vowels.

The Psychological Reality of Phoneme Centers

There is little doubt as to the meaning of phoneme centers as points of maximum agreement in the data from subjects asked to categorize synthetic vowels. Does this lead us to the conclusion that "users of a particular language have a built in pattern which enables them to identify certain acoustic signals as sounds of their language and reject others," (Cohen, 1963, p. 65)? What is the nature of the units of perception of speech? We shall avoid in this study drawing any conclusions concerning these knotty questions. However, it is well to consider at this point how the as yet imperfectly understood nature of the units of speech perception might relate to the type of data we are considering.

In general the various theories of speech perception now current cannot be directly related to data from synthetic speech perception tests of the type we have described. By this we mean that this type of data cannot be used as crucial verification of either of the two most popular theories; i.e. the motor theory and the acoustic theory. By way of clarification, we will describe briefly these two theories.

An acoustic theory of speech perception seeks acoustic cues to match perceptual categories (Fant, 1967; Jakobson, Fant, and Halle, 1952). Since a match between phonemes, or even allophones, and invariant acoustic cues has proved impossible, proponents of an acoustic theory of speech perception look to phonetic features as the basic units.

In contrast to an acoustic theory, there is a motor theory of speech perception (Liberman, 1967; Koshevnikov and Chistovich, 1966). As stated in Kozhevnikov and Chistovich:

> From the motor theory of perception ensues a very important practical deduction: the search for invariant features of phonemes at the acoustic level, which has already involved the expenditure of much time and effort, will hardly yield any positive results. (Kozhevnikov and Chistovich, 1966, p. 204)

Proponents of this theory suggest motor patterns as the basic units of perception, and while accepting the existence of phonetic features at some initial stage of perception, emphasize the organization of perception around the phoneme and the syllable.

Arguments for and against both theories have consumed many scholarly pages (Denes, 1965; Galunov and Chistovich, 1966; Lane, 1965). As in many such discussions, there seems to be a certain amount of semantic confusion as to the claims of each theory (For perhaps the clearest exposition of the arguments, see Galunov and

Chistovich, 1966). Without entering into this controversy, we may ask ourselves in what way the truth of one of these two theories may affect the study. Not in question is the assumption that speakers have some standard pattern in their minds, to which they compare stimuli. As will be discussed later, it can be shown that these patterns are related to the linguistic experience of the subjects. Directly or indirectly, a listener must be able to compare his standard to an acoustic stimulus. This implies, however, nothing as to the acoustical or motor basis of the pattern, nor anything about the size of the unit involved. Nor does the motor theory imply anything as to the relation between articulatory patterns and perceptual patterns. This last point will be discussed further below. In short, both theories explain the ability of subjects to categorize synthetic stimuli.

As concerns the size of the basic unit of perception, suggestions range from feature to allophone to phoneme to syllable. Some experiments give evidence for the existence of perceptual categories suggestive of allophones (Chistovich, Fant, and Serp-Leitao, 1966). The relative importance of features, allophones, phonemes and other units in speech perception is yet to be answered.

Liberman (1967) gives, as evidence for his motor theory of speech perception, data to the effect that vowels are perceived continuously while stops are perceived categorically. As evidence for this he gives data showing that for consonants, latency of response in categorization increases at phoneme boundaries, as does discriminative ability. This he finds less true for vowels. This might suggest to some that cross-over points or phoneme boundaries are less well defined for vowels than for consonants. Further consideration of this point, as well as personal conversation with Philip Lieberman, a proponent of the theory, leads to the conclusion that the non-categoriality of vowels is in no way related to the shape of the cross-over boundary observed in categorization of synthetic stimuli.

Apparently by categorial perception Liberman means the inability to pattern or order stimuli within a category. A stimulus is not heard as the consonant /b/ with a bit of /d/ coloring. Vowels can be heard for example as /i/ with a bit of /e/ coloring. The data which defines categorial from non-categorial perception is based solely on tests of discrimination of the A-B-x variety, where the subject hears two stimuli, A and B, and one of these two is repeated, the subject having to say which of the two he has heard. The non-categorial nature of vowels is thus not reflected in the categorization tests we are using.

As a final word on the nature of speech perception, and the psychological reality of various units, we may say that the subject is undoubtedly far more complex than any current theory allows for. As D. B. Fry expresses it:

> I suggest we should not refer to 'the motor theory of speech perception,' nor even to 'a motor theory,' because speech can be perceived in a number of ways. Perception of speech at any time will be the result of a number of different factors, and what we are trying to find out is what weight should be given to the effect of motor memories in given circumstances. (reply to Denes, 1965, in Proc. of the Fifth International Congress of Phonetic Sciences, p. 258)

C. THE RELATION BETWEEN PERCEPTION OF SYNTHETIC VOWELS AND PERCEPTION OF NATURAL LANGUAGE

As described above, experiments in vowel perception indicate that subjects behave as if they have learned a set of standard vowel patterns, or perceptual categories. These categories may be reasonably satisfactorily specified by areas on an F1/F2 chart (see Appendix B-1). We will now ask to what degree these categories may be related to the perception of natural language. A number of studies indicate that such a relationship exists.

Perception and Native Language

Several experiments have used natural speech sounds spoken and perceived by groups of speakers of each of several languages (Singh, 1966; Singh and Black, 1966). Singh and Black had speakers of Hindi, English, Arabic, and Japanese record 26 consonants spoken inter-vocalically. Subjects from each of the four language groups were then asked to identify the consonants. Results were presented in a confusion matrix, with rows showing intended consonants and columns showing perceived consonants. The study showed that there was greater confusion across languages than among speakers of the same language. Confusion was not random, but predictable according to differences in the phonology of the various languages. In particular, perception of voicing and aspiration reflected language differences.

Lisker and Abramson (Lisker and Abramson, 1964) reported a cross-language study of voicing, using a series of synthesized consonants in which the voice onset time varied from extreme prevoicing to late voicing, so that consonant stimuli ranged from voiced to unvoiced. They found that speakers of different languages divided the series between voiced and unvoiced at different points, that is with different cross-over values. (This case also points out how use of synthetic stimuli produces information as to cross-over points which is not possible with the natural language sounds described previously, since close control of voice onset would not be possible with natural sounds.) Delattre (Delattre, 1964) made a study of vowels as perceived by speakers of English, French, German, and Spanish, using synthetic stimuli, and found that preferred vowels (i.e. phoneme centers) varied in formant frequency for each of the four languages. Scholes (Scholes, 1967) did a similar study of "Phoneme Categorization of Synthetic Stimuli by Speakers of Japanese, Spanish, Persian, and American." Studies such as the above cited prove conclusively that performance in categorizing synthetic stimuli reflects in a rather stable way the native language of the subjects.

Perception and Dialect

What is true for different languages may well be true for different dialects of the same language. This idea is indeed suggested by Obrecht (Obrecht, 1965) and is a major hypothesis of the present study. Obrecht found that speakers of different dialects of Arabic gave different categorizations of some synthetic speech sounds.

The possibility of dialectal variation in categorization of speech sounds has not been thoroughly investigated, and there is evidence both for and against such a relation. Scholes has examined his data for Greek subjects and finds no indication of dialectal variation.[1] He suggests that the Arabic dialects considered by Obrecht are perhaps so dissimilar as to be separate languages, and doubts whether perception tests would be sufficiently sensitive to small differences in formant frequency, so as to be a useful indicator of dialect. Scholes however did not specifically investigate the possibility of dialect influences, and his evidence against it is not conclusive.

---

[1] "It would appear, then, that the experimental procedure described here is not suitable for dialect discrimination or identification." Scholes, unpublished paper.

Peterson and Barney (Peterson and Barney, 1953, p. 177ff) and Ladefoged and Broadbent (Lodefoged and Broadbent, 1957, p. 103) give evidence for the perception of synthetic vowels being influenced by English dialect. Leon and Nehmi (1967) asked French subjects to listen to words pronounced by French Canadians, and observed confusions consistent with differences in dialect. They find for example that all 18 listeners mis-heard Canadian "Quelle aire" as "quelle heure," indicating a correlation with the fact that Canadian /E/ resembles French /oe/. The authors present their data only as a preliminary study. Lindner (Lindner, 1967) made a thorough study of variation in perception of synthetic vowels for 100 German subjects. He sees a correlation between intra-subject variation and dialect, but is unclear as to how such a correlation is to be explained in terms of the dialects in question.

The Relationship Between Articulation and Perception

When considering the possible influence of dialect on perception, it is important to consider the as yet imperfectly understood relation between an individual's perception of a set of speech sounds, and his articulation of these same sounds. Most speakers of a language have a limited experience in perception of other languages. When comparing subjects who have different language backgrounds, it is never clear whether in perceiving synthetic vowels they are perceiving according to how they speak, or how they are accustomed to listening. Speakers of a particular dialect of English, on the contrary, even if they speak only that dialect, must certainly have considerable experience in perceiving speech of a number of other dialects. The degree to which speakers of a given dialect perceive according to that same dialect may depend on the extent to which perception depends on articulation. We should hasten to add that the motor theory of speech perception as we understand it, in spite of its emphasis on articulatory patterns, implies no decision on this question. The motor theory suggests that the basic unit of perception of speech sounds is a motor nerve pattern or command. It is still quite possible that a variety of acoustic patterns can be mapped on to one such unit. That is to say, it is possible that a single speaker, with a single basic motor pattern for a given sound, learns to associate with this unit a number of dialectal variations.

There have been a number of experimental investigations into the relationship between speech sound discrimination and articulation, none of them conclusive. Griffith (Griffith, 1952) found no correlation between ability to articulate sounds well and ability to perceive the same sounds correctly. Aungst and Frick (Aungst and Frick, 1964) review work done in speech pathology relating to this problem and conclude that "research to determine the relationships between auditory discrimination ability and articulatory ability has yielded equivocal findings." (p. 76) The authors point out that to the degree that auditory feed back is needed for proper articulation, articulation and perception must be related. They make the distinction between "internal" perception, the monitoring of ones own speech, and "interpersonal" perception, i.e. listening to others. They then present results of an experiment showing that for the phoneme /r/, internal discriminative ability correlates with articulatory ability, while intrapersonal discriminatory ability does not. To the extent that the Aungst and Frick study is relevant to the perception of synthetic vowels, it leads to the conclusion that there is no direct relation between articulation and perception of sounds other than in monitoring one's own speech. However, the Aungst and Frick study uses different methods and terminology from our own and is of limited relevance.

Menyuk (Menyuk, 1967) found that children do not categorize vowels differently from adults. Since it is known from spectrographic analysis that children's vowel formants are consistently different from those of adults, she concludes that there is no relation between perception and articulation.

One experiment that does indicate an interdependence between articulation and perception is Lorge (Lorge, 1967). She compared perception of synthesized initial /d/-/t/ with articulation of these sounds by four subjects, in terms of onset of voicing. The author concludes that the two subjects who showed least difficulty in discriminating /d/-/t/ were found to have a greater proportion of articulated /d/ responses with considerable pre-voicing, than did the two subjects who performed less well in the perception test. The results are inconclusive due to the small number of subjects, the fact that in none of the subjects was there any overlap of onset time for /d/-/t/, and due to the generally preliminary nature of the data.

We may conclude that articulation of speech and perception of synthetic speech sounds do not appear to be interdependent. If this is true then we should find in our tests that subject responses reflect more the dialectal environment of the subject than they reflect his own dialect. If it is observed that a subject with dialect A, who has been living several years in an area of dialect B, perceives according to what would be expected of dialect B, we may regard this as further evidence that perception and articulation are independent of each other. If on the other hand a subject persists in perceiving in a manner directly related to his own dialect, regardless of where he lives, we may conclude the opposite.

Individual Variation in Vowel Perception: Performance and Reliability

In order to correlate group results on categorization tests with dialect, our data must be reliable. Intra-subject variation in categorization must not be so great as to obscure group tendencies beyond recovery. Intra-subject variation may be divided into two types: 1) inconsistency within one and the same test, and 2: variability over time. Absence of inconsistencies we shall term here "performance." Poor performance on a test will be indicated by indistinct or non-existent phoneme boundaries. Such performance is an indication of subject uncertainty in performing the task, and would seem to depend on the ability of subjects to associate the stimuli with some standard pattern.

Changes in response to the same stimuli upon different administrations of the same test, we shall term "variability." It is possible that subjects will show good performance but that their phoneme boundaries shift with time and circumstances.

Laver (Laver, 1965) studied the performance of trained linguists in associating synthetic vowels with position on the traditional vowel triangle. He found that a given subject did vary over time, in the place on the triangle to which they assigned the stimuli. The degree of variation depended on the particular subject, and varied from subject to subject. Laver gives the following comment on the causes of variability:

> ..... the theoretical causes of variability must also wait for a further detailed discussion. However one or two hypotheses may be briefly noted...... Firstly, if, for instance, a motor theory of speech perception (Liberman, et al., 1962) is valid, and applicable to vowels, then the neural complexes by reference to which vowel-perception would be said to be mediated, may not be as stable as orthodox cardinal vowel theory would demand. Secondly, the physiological and physical aspects of the situation may vary slightly, and the combination of these and psychological factors might influence the ability of a particular subject to concentrate. Thirdly, exposure to vowels of one particular general quality may act as an adaption-level factor in individual cases. Lastly, the use of synthetic rather than genuine human material may, in some way yet to be investigated, affect the ability to locate vowels on a chart. (Laver, 1965, p. 120)

Lindner (Lindner, 1966; 1967) made an extensive study of consistency in perception of 10 synthetic vowels by Germans. He tested 57 subjects twice, three months apart, and found that on the average, 73% of the judgements were unchanged. There were considerable inter-subject differences in reliability, individual percentages of consistency ranging from 55% to as high as 85%. That is to say that at least one subject made the same categorizations for 85% of the items, three months later.

The degree of intra-subject variation will certainly depend on the nature of the stimuli, and their naturalness. It will be crucial to our study that intra-subject variability not obscure group tendencies correlating with dialect. A study of performance and reliability will be one aim of this study.

II. METHOD OF INVESTIGATION

A. AN OVER-VIEW OF THE EXPERIMENTS

In the present study we shall administer a series of synthetic vowel perception tests to speakers of various dialects of American and Canadian English. We shall inquire as to:

1) Individual performance;
2) Group performance and reliability;
3) Group tendencies, and especially those tendencies correlating with features of dialect.

Our major hypothesis is that categorization of synthetic vowels by speakers of a dialect will reflect differences in quality of these same vowels as articulated by speakers of this dialect.

As explained above, we are assuming that the categorization tests reflect, at least indirectly, perceptual categories used in perception of natural language. We are further assuming that these categories can be satisfactorily described on an F1/F2 chart.

The vowels to be studied are the vowels of the words: bead, bid, bed, bayed, bad, bud, bawd, box, bowed, book, booed. We shall not consider the diphthongs of the words: boy, house, buy.

The dialects considered are those of subjects living in Ft. Erie, Canada (Ont.), Buffalo, N. Y. (Buf.), and Rochester, N. Y. (Roch.), Culpeper, Va. (Va.), and Oxford, Mass. (Mass.).

Group tendencies will be analysed in terms of the number of phoneme boundaries and centers observed, and the positions of these on the F1/F2 chart. The existence or non-existence of phoneme centers and boundaries will be compared to the existence or non-existence of phoneme contrasts in the dialects of subjects producing the data. Position of phoneme borders will be compared to the phonetic values of phonemes in these dialects.

The present study is unique in being an extensive study of intra-subject and group variation in the perception of synthetic vowels by speakers of English. It is also the only investigation known to the author, concerning the relation between perception of synthetic vowels and English dialect. We have attempted to improve on some previous studies, particularly early ones at the Haskins Labs, since our stimuli are considerably more speech-like, or human, than those produced by the Pattern Playback synthesizer at Haskins. Whereas many previous studies have been made with pitch and amplitude at a steady value, we shall use more natural sounding pitch and amplitude functions. It is likely that the addition of F3, F4, and F5 as invariables increases the speech quality of our stimuli as well. Carpenter and Morton (Morton and Carpenter, 1963), noting the importance of naturalness hypothesize that:

> ..... if a sound lacks naturalness, it is possible that we initially reject it as being a vowel and only then consider the vowel color in abstract..... In other words we could postulate an interaction between the separable attributes of naturalness and vowel color which ought to be considered...... (p. 214)

Our study will differ from previous studies, including Scholes (Scholes, 1967), in that the values of the stimuli will be more closely spaced together on the F1/F2 chart. Scholes doubts whether the relatively small phonological differences among dialects can be detected in categorization tests.[2] Whereas Scholes chose intervals of 100Hz or greater for formant frequencies, we will use intervals of 50Hz or less.

B. PRODUCTION OF THE SYNTHETIC VOWEL STIMULI

249 stimuli were produced on EVA, a commercially manufactured speech synthesizer.[3] Parameters which were used are pitch (F0), amplitude (A0),

---

[2] cf. Scholes, ibid, ft. nt. 1.

[3] EVA, Mark II, is produced by Melpar Inc., Falls Church, Va., and operated by the Phonetics Laboratory, University of Rochester, under the direction of Dr. Dean Obrecht.

and peak frequency of four resonator circuits (connected in series), which specify four formants (F1, F2, F3, F4). Also used was an additional formant (F5), the peak frequency of which is controlled by a dial with five settings. F5 is of relatively little importance in determining the quality of the stimuli. EVA has nasality and noise parameters which were not used. Each of the parameters except for F5 can be controlled manually or by an ink-drawn pattern. If controlled manually, the value of the parameter is predetermined by a dial setting, and remains constant for a given stimulus. In the second case a pattern is drawn in electrically conducting ink and scanned by a roller, in which case the value of the parameter may change over time as programmed by the ink-drawn pattern. For the production of stimuli used in these tests, F0 and A0 were programmed for patterns designed to give a vowel sound resembling human utterances. All other parameters were controlled manually and remained constant for a given stimulus. (Settings for all parameters are given in Appendix A-1.)

Since the true values of F1 and F2 do not correspond to readings on the appropriate dials, direct measurements were made of peak frequencies of F1 and F2 for each stimulus. These measurements were not made of the stimuli themselves, but of "calibration tones" recorded along with each stimulus as it was produced with all settings identical to those of the stimulus, except that A0 and F0 were switched to manual, constant settings, and F0 was lowered to 60Hz. The lower and constant F0 made reading of formant peaks easier. A comparison of a cross-section of an actual stimulus with a cross-section of the corresponding callibration tone showed that the measured peak frequencies of the two cross-sections were identical.

Each pass of the synthesizer roller over the ink pattern produced two identical stimuli, .3 seconds in duration, separated by a one-second interval.

It is more accurate to consider F1 and F2 values as relative rather than absolute, since there is considerable error in these values. (Variation in EVA production for a given setting, and variation due to recording was negligible.) As relative values, the measurements are accurate to at least ± 20Hz, this being the limit of accuracy due to personal error in reading peaks on a spectrogram with the 3000Hz scale used. There was a larger, systematic error due to the non-linearity of the sonagraph scale, as compared to an audio-oscillator of established accuracy. This error was at some frequencies as great as 60Hz, but since the error was invariable over time, it did not affect the relative position of the formants as measured.

C. THE COMPOSITION OF THE TESTS

The 249 stimulus items were recorded and then copied and formant frequencies measured as described above.[4] From these stimulus items three series of tests were made, in the following manner. The copied stimulus items were cut out and spliced together in random orders, with a .5 second interval between items. (A stimulus item is two identical stimuli separated by a one second interval.) These were then recopied, so that the final test was a copy of a copy. Care was taken not to lose fidelity and while there was probably some loss, vowel quality was not noticeably affected. The final copy had a 10 second pause every fifth item, at which time the number of the next item (6, 11, 16, etc.) was given. The number of items in a test ranged from 28 to 36. Average test time was five minutes.

There were three series of tests. Series I (Tests IA, IB) were composed of stimuli in the quality range of the phonemes /i/, /e/, /E/, /ae/, /a/, /ɔ/, /o/, /u/. Test IA was designed as a pre-test to establish roughly the phoneme centers for the dialects studied. This test was composed of stimuli covering the entire range of stimuli produced. Test IB, while designed to test phoneme centers for the same eight phonemes as IA, was composed of stimuli from areas of the F1/F2 chart shown to be relevant to these phonemes, as a result of pretests with IA.

---

[4] Recording and copying were done with a Magnecord 1022 and an Ampex 601.

Series II (Tests IIA, IIB, IIC) were composed by reference to phoneme centers determined by the pre-test IA. In these tests stimuli were limited to areas of the F1/F2 chart relevant to no more than two phonemes, and subjects were asked to divide the items into two categories, as opposed to the eight or more categories of Series I. In Series II 30 to 33 different stimuli were presented for each test, from the relevant F1/F2 area.

Series III (Test IIIC) resembled Series II in that subjects were asked to divide the items into one of two categories. However, Test IIIC was composed of eight stimuli, presented four times per test. Tests of the II series offered subjects a greater choice of stimuli, but gave no measure of variability. Test IIIC gave an indication of variability in that each stimulus appeared four times in the same test. Appendix B gives the composition of all tests. Note particularly Appendix B-1, which explains the F1/F2 facsimile charts which will be used to present data.

D. ADMINISTRATION OF THE TESTS

All tests were administered using a Wollensak 1500S tape recorder, with tone control at "Hi-Fi" and volume set to suit testing conditions. The tests were administered on location, usually in high school classrooms. All subjects of a given group were tested together, as a group.

Whenever possible the tests were administered by a teacher or other adult familiar to the group of subjects, with the experimenter observing and helping out. After an informal introduction to the testing procedure, the subjects were read brief instructions. After the fifth item of each test, the recorder was stopped, and subjects were allowed to ask any questions they wished concerning procedure. (This was done only as needed, as determined by the reaction of the subjects.)

Each subject was also asked to fill out a questionnaire which was used to establish to what degree he was native to the region he was living in when tested. In some cases the voices of the subjects were recorded.

The questionnaire, instructions, and sample answer sheets appear in Appendix C. There were two versions of Test IB. In the "open-ended" version, subjects were asked to write "any one syllable English word." In the forced choice version, they were asked to choose one of 9 words given on the answer sheet. These nine words were chosen to represent the 8 major response categories (see page 36), with the addition of the word "tar," representing the category /ar/ which is a separate phoneme in dialects of Eastern Massachusetts.

E. DESCRIPTION OF THE SUBJECTS

The main body of our data is based on the testing of 196 subjects from four dialect regions.[5] The regions selected were: Western New York State; Virginia; Ontario, Canada; and Massachusetts. In most cases the groups tested were high school freshmen and sophomores. In Massachusetts and Rochester, N. Y., Girl Scouts and Campfire Girls were used as subjects. Ages ranged from 13 to 16, with most subjects being 14 or 15 years old. (At times later in the text reference will be made to subjects not described in this section. These cases will be duly noted in the text.)

Appendix D gives a description of the sex, I. Q., and "dialect background" of the subjects. For most of the groups, no direct check was made of the spoken dialect of subjects, it being assumed that they used the dialect features reported to be typical of their region. Information from the questionnaire was used as an indirect check of dialect. Subjects were classified as to background: Native, if both parents and the subject were born in the region, and the subject was raised in the same region; Non-native, if the subject was born outside the region. All groups tested seemed to be representative of their region.

---

[5] Thanks are due to the teachers and students of Lafayette High School, Buffalo, N. Y.; Ft. Erie High School, Ft. Erie, Ontario, Canada; and Culpeper High School, Culpeper, Va., to the leaders and members of the Girl Scout Troup of St. Monica's Parish, Rochester, N. Y., and to the Girl Scouts and Campfire Girls of Oxford, Mass., who served as subjects.

Tighter controls were kept on the dialects of the Massachusetts and Rochester, N. Y. groups. Firstly, we may assume that these groups represented a more homogeneous socio-linguistic grouping than did the groups from high schools. We say this under the assumption that a group of girls such as Scouts or Campfire Girls is a voluntary association, determined by social ties, whereas a high school class is less determined by social ties. Secondly, as a further control of the dialect of subjects in these last two groups, recordings were made of each subject reading a written paragraph. From these recordings the experimenter established that all subjects had typical features of the dialect of their region.

The following is a description of the groups tested (See also Appendix D). The abbreviated names given here will be used henceforth.

1. WESTERN NEW YORK STATE REGION

Group Buffalo A (Buf. A) - 22 subjects from a freshman English class at Lafayette High School in Buffalo, N. Y. Tests IA, IB, and one week later, Tests IIA, B, C were administered.

Group Buffalo B (Buf. B) - 16 subjects from a freshman English class at Lafayette High School. Tests IB, IIA, B, C were administered.

Group Clarence (Clar.) - 11 subjects from Clarence High School, Clarence, N. Y. (near Buffalo). Tests IB, IIA were administered twice at a one week interval.

Group Rochester A (Roch. A) - 21 subjects from St. Monica's Parish Girl Scout troup, Rochester, N. Y. Tests IB, IIA, B, IIIC were administered twice at an interval of one week.

2. ONTARIO, CANADA REGION

Group Ontario A (Ont. A) - 28 subjects from a freshman French class at Ft. Erie High School, Ft. Erie, Ontario, Canada (just across the river from Buffalo). Tests IA, IB, and one week later Tests IIA, C were administered.

Group Ontario B (Ont. B) - 24 subjects from a freshman English class at Ft. Erie High School. Tests IB, IIA, C were administered.

3. VIRGINIA REGION

Group Virginia A (Va. A) - 30 students from a freshman

English class at Culpeper County High School, Culpeper, Va. Tests IA, B, IIA, B, C were administered.

Group Virginia B (Va. B) - 23 students from a freshman English class at Culpeper County High School. Tests IA, B, IIA, B, C, were administered.

4. MASSACHUSETTS REGION

Group Massachusetts (Mass.) - 21 subjects from the combined Junior and Senior Girl Scouts and Campfire Girls of Oxford, Mass., near Worcester. Tests IB, IIA, B, IIIC were administered twice at an interval of several weeks.

F. DESCRIPTION OF THE DATA

Raw data for all tests consisted of answer sheets with words, for series I, and checks, for series II, III, as responses. These answer sheets were encoded into computer readable data (Appendix E-1). In order to reduce the answer sheets to computer readable data, the responses were encoded in thirteen categories; eleven phoneme categories, other responses, and no response. Table I explains how this was done. The data was tabulated by computer (Appendix E-2), and for each subject the computer printed out the responses of each subject to each test, in an individual F1/F2 facsimile format (Appendix E-3). The tabulated group results were converted to percentages and displayed on group result F1/F2 facsimiles. (Fig. A shows a sample of group results.) These group result F1/F2

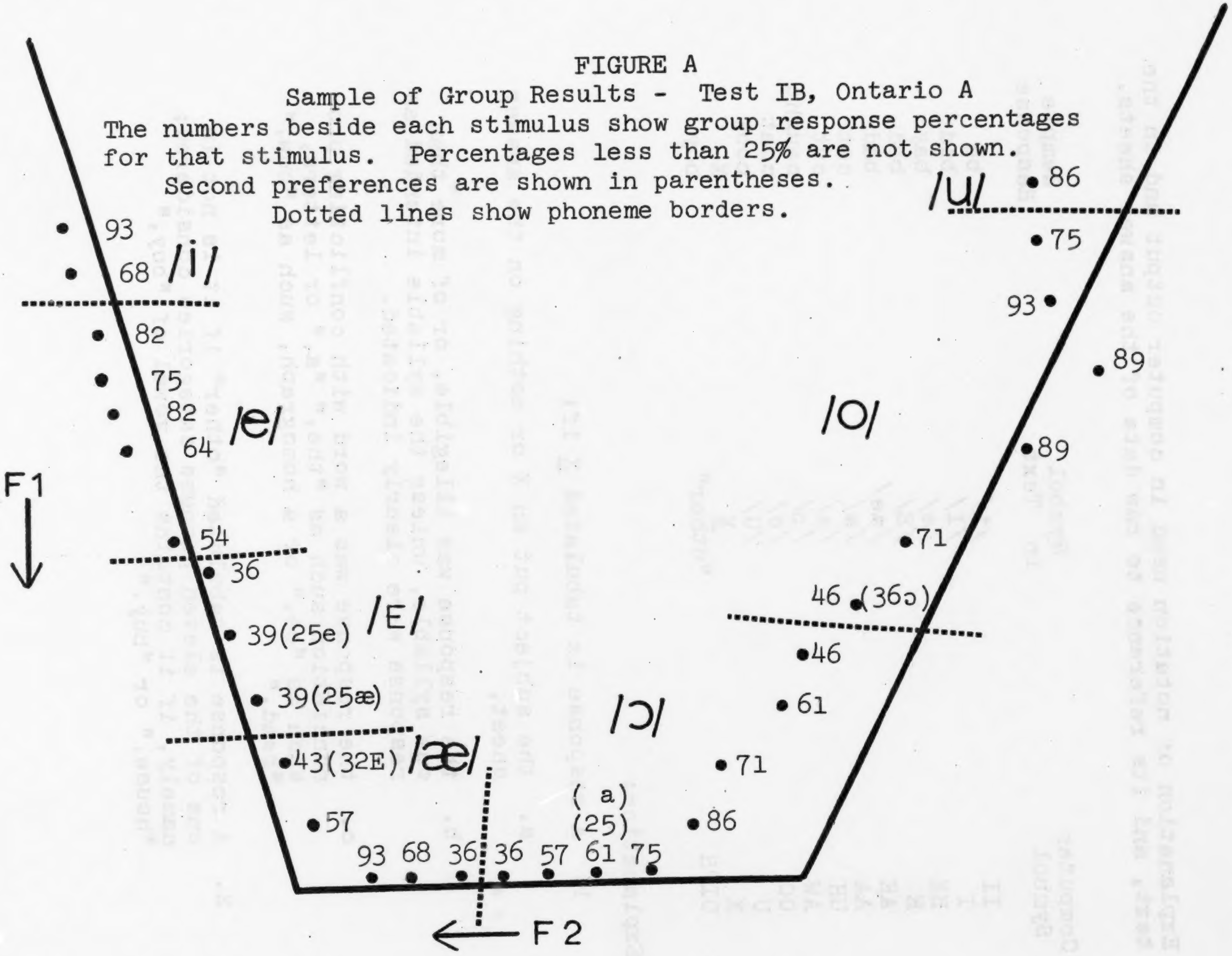


FIGURE A

Sample of Group Results - Test IB, Ontario A

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown. Second preferences are shown in parentheses. Dotted lines show phoneme borders.

TABLE I

Explanation of notation used in computer output and in the text, and its reference to raw data of the answer sheets.

Computer Symbol	Symbol in Text	Sample Response
II	/i/	be
I	/I/	bid
EE	/e/	bay
E	/E/	bed
AE	/ae/	bad
AA	/a/	cot
UH	/ʌ/	cut
AW	/ɔ/	caught
OO	/o/	coat
U	/U/	book
X	X	X
DIPH	"other"	boy

Explanation:

1. A response is tabulated X if:

   a. the subject put an X or nothing on the answer sheet,

   b. the response was illegible, or of more than one syllable, unless the syllable intended as response were clearly indicated.

   c. the response was a word with conflicting pronunciation such as "the," "a," or letters, such as "ee," or a homograph, such as "bow," "read."

2. A response is tabulated "other" if it is not in one of the eleven phoneme categories considered; namely, if it contains the vowel of "boy," "house," or "buy."

facsimiles are the form in which the raw data are presented in this thesis.

In Test IA, and in the open-ended version of Test IB, subjects were asked to write any one-syllable English word. With such a wide variety of possible responses, there was some unavoidable loss of information in reducing the answer sheets to computer readable data. The number of responses lost in categories *X* and "other response" was small enough not to necessitate attempts to recover what had been intended by the subject. In any case such recovery would have been generally impossible.

One further problem which caused some loss of information (in the open-ended version of the tests) was the appearance of responses such as the words "aunt," "fog," "caught," "cot," and others which vary dialectally as to the phoneme category into which they should be placed. Since it was not possible to know what the subjects intended, such words were recorded according to the experimenter's dialect. It should be noted that such a procedure could not produce any false tendencies in the data, but only result in a loss of information about such tendencies. Fortunately such words were uncommon in the data. In any case, this weakness of the open-ended version was avoided in the forced choice version of Test IB, which allowed only nine category words to be used as responses.

III. RESULTS

A. MAJOR RESPONSE CATEGORIES

Before asking subjects to categorize stimuli into a limited number of phoneme categories, (as in the forced choice version of Test IB), it was preferable to find out experimentally what categories were preferred responses to the stimuli (major response categories). Subjects taking Test IA and the open-ended version of Test IB were asked for "any one syllable English word." Data from these tests were used to establish major response categories. These categories then became the categories used in the forced choice version of Test IB.

Table II gives the percentages of responses in twelve categories for the two tests (X responses are not shown). Eight categories (/I/,/e/,/E/,/ae/,/a/,/c/,/o/, /u/) have the highest percentages and are therefore judged major response categories, for the subjects who took Test IA and the open-ended version of Test IB. These eight categories accounted for from 60% to over 80% of all responses, depending on the group.

B. COMMON PHONEME CENTERS

Common phoneme centers were determined using the data from Test IA. Phoneme centers for each dialect were established in terms of the stimuli preferred for the

TABLE II

Major Response Categories for Tests IA and IB, open-ended version.

Test IA

	II	I	EE	E	AE	AA	UH	AW	OO	U	UU	DIPH
Buf.A	17	1	3	1	13	6	3	10	12	0	8	3
Ont.A	13	1	11	7	6	2	1	11	16	1	7	1
Va.A	12	2	8	11	9	3	3	13	13	1	9	2
Va.B	15	2	5	12	8	4	2	15	10	4	9	2

Test IB

	II	I	EE	E	AE	AA	UH	AW	OO	U	UU	DIPH
Buf.A	8	3	8	5	12	6	9	14	12	0	8	3
Buf.B	7	2	6	3	20	5	1	15	13	0	5	2
Ont.A	6	0	15	8	13	3	1	19	16	1	4	0
Ont.B	7	0	12	3	16	6	1	13	15	0	5	5
Va.A	7	2	9	9	14	11	2	13	12	0	7	2
Va.B	8	3	5	14	10	12	1	16	10	1	7	3

Explanation:

Categories I, UH, U, DIPH have low percentages of response for all groups.

The remaining categories are judged to be major response categories, on the basis of these data.

Note:

1. These data are taken from the open-ended administration of the tests, in which subjects were asked to respond with "any one syllable English word."
2. Numbers are percentages rounded to the nearest whole percent.

various phoneme categories. Stimuli which elicited the same category for all dialect groups were considered to be common phoneme centers. The determination of these common phoneme centers served as a check on the other Tests (IB, IIA, B, C, IIIA, B, C) in that we could be relatively certain that these tests contained stimuli appropriate to the response category demanded, for all dialects.

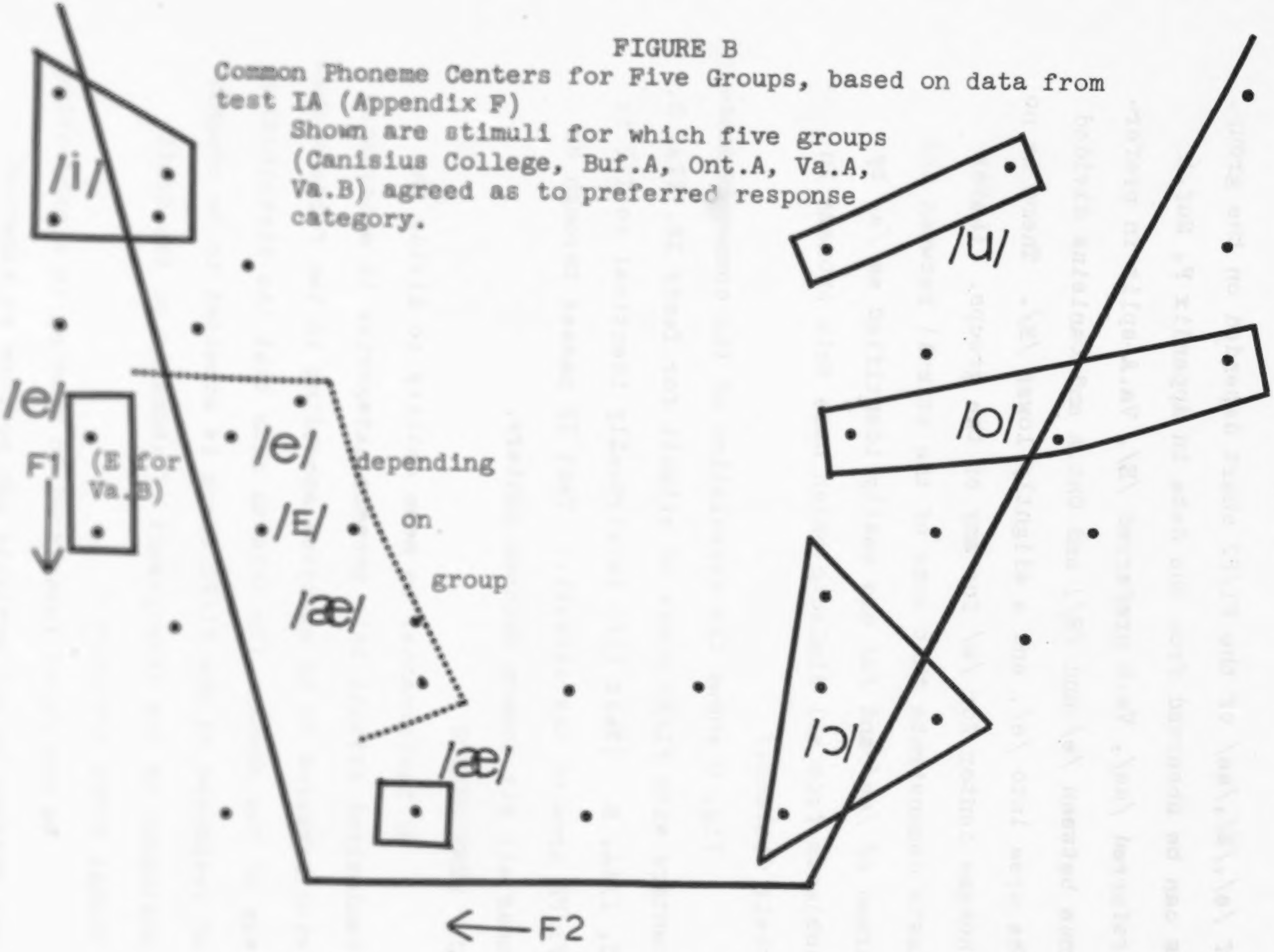
Ideally Test IA would have been administered to all dialect groups, and then on the basis of the data, succeeding tests would have been constructed, but this was not feasible. The actual procedure was to administer Test IA to a group of 50 subjects, (college students from Canisius College, Buffalo, N. Y.) and on the basis of their responses, to construct the succeeding tests. As a check that the tests were composed of stimuli appropriate to the category of response demanded, Test IA was administered to Ontario, Virginia, and Buffalo groups, at the same time as the other tests. Results are given in Appendix F.

Fig. B shows common phoneme centers as determined from the data in Appendix F. There are areas of agreement for the /i/,/e/,/ae/,/o/,/e/, and /u/. Percentages of agreement for the /a/ center were not high, and Va.B preferred /E/ for this stimulus. Preferences in the area

FIGURE B

Common Phoneme Centers for Five Groups, based on data from test IA (Appendix F)

Shown are stimuli for which five groups (Canisius College, Buf.A, Ont.A, Va.A, Va.B) agreed as to preferred response category.



of /e/,/E/,/ae/ of the F1/F2 chart depended on the group. As can be observed from the data in Appendix F, Buf.A preferred /ae/, Va.B preferred /E/, Va.A split in preference between /e/ and /E/, and Ont.A and Canisius divided the area into /e/, and a slightly lower /E/. There is no phoneme center for /a/ for any of the groups. (Later tests demonstrate that some of the stimuli between the areas of /ae/ and /ɔ/ are easily identified as /a/ by subjects from all dialects which have this phoneme in their dialect.)

Fig. C shows the correlation of the common phoneme centers with F1/F2 areas of stimuli for Tests IB, IIA, B, C, IIIA, B. (Test IIIC is virtually identical to IIC in F1/F2 area of the stimuli.) Test IB passes through or near all six common phoneme centers.

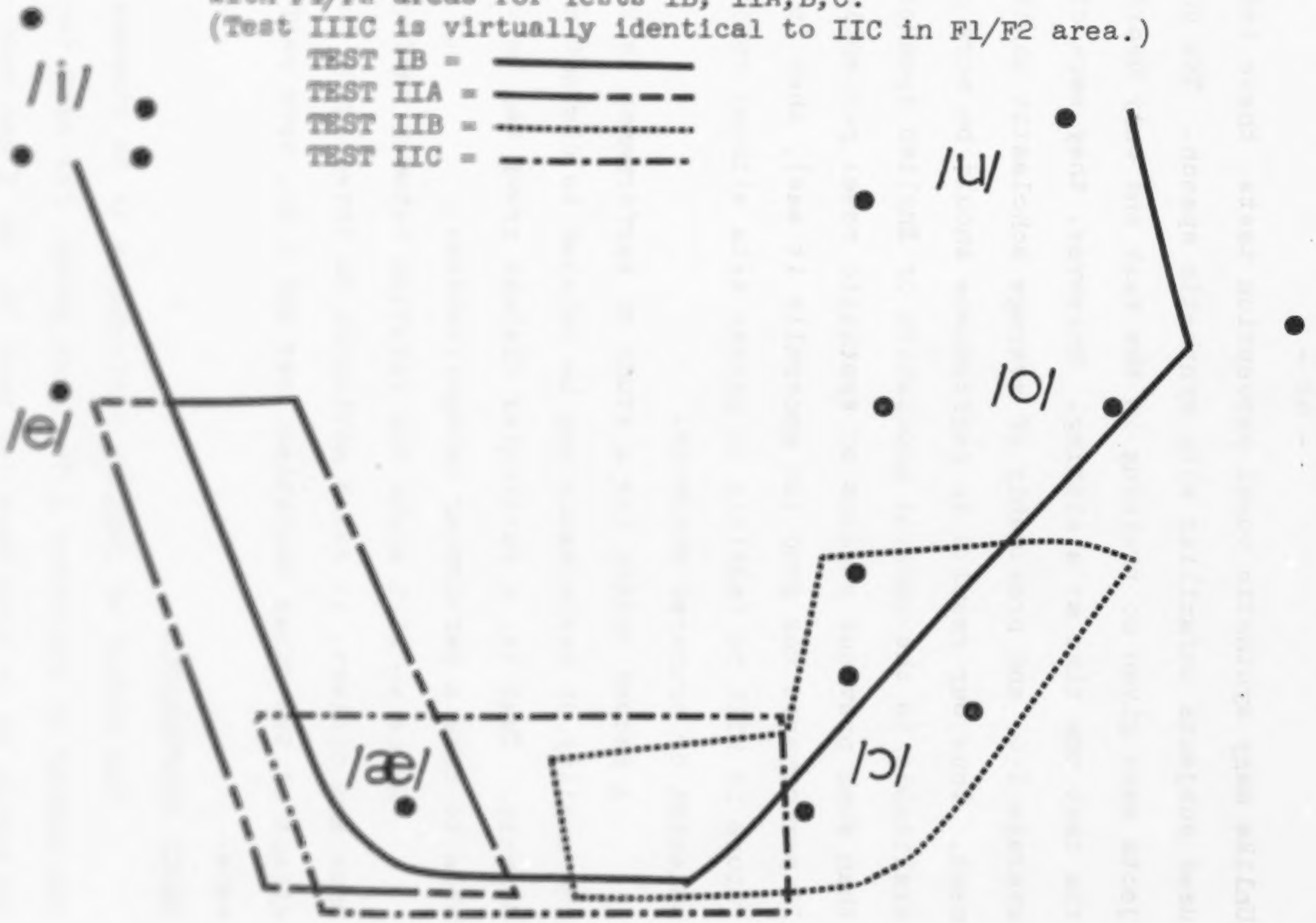
C. PERFORMANCE

By performance, we mean ability to divide the randomized stimuli into phoneme categories in accordance with patterns to be expected according to two formant theory of the vowel. (By this we mean that the distribution of responses on the F1/F2 chart is expected to be roughly analogous to the arrangement of phonemes on the traditional vowel triangle.)

We are first interested in knowing in a subjective way whether or not subjects can perform as expected.

FIGURE C

Correlation of 6 Common Phoneme Centers (cf. Fig. A), with F1/F2 areas for Tests IB, IIA,B,C. (Test IIIC is virtually identical to IIC in F1/F2 area.)



Unlike many synthetic vowel perception tests, these tests used subjects unfamiliar with synthetic speech. The subjects were given no training in the task and only heard the test one time at a sitting. Moreover, they were of average I.Q. and presumably of average scholastic achievement. Thus our results in performance should be more generalizable to the general population of English speakers than some previous studies of synthetic vowel perception. If performance was good (as generally it was), then in the future it will be feasible to gather data without specially selected or prepared subjects.

A second reason for a study of performance is that the quality of performance may be related to dialect membership. That is, a particular dialect group may be unable to make a particular categorization.

Since we will argue the relation between performance and dialect, it seems advisable to investigate the effect of two other variables, sex and I.Q., upon performance.

GROUP PERFORMANCE

One method of judging performance is to compare the number of responses X for each group, for each test (although it is true that for Test IB, an X can mean an illegible or uninterpretable response, as well as failure to respond. cf. Table I). For Tests IIA, B, C the number of responses X per group per test was never greater than 3% and in most cases less than 1%. For Test IB the percentages were considerably higher. Table III gives the percentages of response X for Test IB for nine groups. The striking reduction in percentages for Rochester and Massachusetts is most probably due to the fact that these tests were administered with the multiple choice version of Test IB, in which the subjects had to choose among nine category words printed at the top of the answer sheet. Thus the high percentages of X for other groups is probably due to testing procedures, and not to the inability of subjects to associate the stimuli with English words. (The fact that Mass. and Roch. were all female may have also been a factor in performance. cf. below, "Sex and Performance.")

TABLE III

Performance on Test IB in terms of percentage of response X.

	Buf.A	Buf.B	Ont.A	Ont.B	Va.A	Va.B
responses X in percent.	14.0	20.0	13.0	16.0	12.0	9.0
	Clarence		Rochester		Massachusetts	
	1st	2nd	1st	2nd	1st	2nd
responses X in percent	12.0	6.0	0.6	0.4	0.3	0.0

A second method of judging performance is in terms of phoneme centers. Table IV shows the response percentages for seven phoneme centers (i.e. stimuli with highest response percentages), for Test IB. Since most groups did not have both /e/ and /E/ phoneme centers, percentages for this category are based on the highest of the two categories, for each group. These percentages are based on data given in Appendix G. There are high response percentages (i.e. good performance) for /i/,/o/, /u/. Phoneme centers for /ae/ and /E/e/ show percentages generally over 40%. /a/ has the lowest percentages. In general all groups show four categories with high percentages, and many show more. An examination of Appendix G shows that all groups give patterns consistent with what might be expected according to two formant vowel theory.

INDIVIDUAL PERFORMANCE

In order to describe the performance of an individual for a particular test, we divised a method of judging good performance. In general this method consisted of establishing a standard criterion for determining whether a phoneme boundary could be drawn between responses as displayed on the computer-produced F1/F2 facsimile format (Appendix E-3). For Tests IIA, B, C, the criterion was that if such a line could be drawn, separating one category from the other, and no more than two responses were

TABLE IV

Response percentages for phoneme centers for Test IB

	/i/	/e/E/	/ae/	/a/	/ɔ/	/o/	/u/
Buf.A	70	35	45	30	60	80	75
Buf.B	87	59	81	31	75	87	87
Roch. 1st.	94	44	67	56	25	83	89
2nd.	94	56	72	83	25	83	100
Ont.A	93	89	93	25	86	89	86
Ont.B	79	75	95	37	70	83	83
Va.A	76	43	70	53	63	80	93
Va.B	73	65	52	52	65	69	78
Mass. 1st.	85	75	100	75	55	80	85
2nd	75	65	85	100	40	85	95

Explanation:

A phoneme center is a stimulus with highest percentage of response for a given category. High percentages are considered here as a measure of good performance.

Percentages in the category /e/E/ are the response percentages either of /e/ or of /E/, depending on which category was higher for a particular group.

displaced from the area of their category, then the pattern was judged good. Test IB required a less precise criterion. Lines were drawn separating one category from the next, on the F1/F2 facsimile. If four or more such lines (phoneme boundary) could be drawn, and there were no more than five responses X, the pattern was judged good.

Table V shows percentages of good patterns per group per test. Percentages for Test IIIC are with one exception over 80%. Performance on Test IIB was poor, with percentages below 53%. Performance on Test IIA varied a great deal from group to group, with percentages ranging from 78% to 0.0%. Performance on the Test IB generally showed percentages better than 50%. It will be noticed that for Test IB, Mass. and Roch. performed the best of all the groups.

PERFORMANCE AND SEX

Table VI shows the percentages of good patterns for male and for female subjects, for four groups. (These were the groups for which sex of subjects was known, and for which the numbers of male and female subjects were similar.) Since performance on Test IIB was extremely poor, and on Test IIC extremely good, these tests were not considered. As is evident from Table VI, females perform better than males in eight out of ten

TABLE V

Percentages of good patterns per group per test

	IB	IIA	IIB	IIC
Buf.A	60	10	23	76
Buf.B	50	12	-	81
Roch.1st	78	0.0	26	-
" 2nd	78	11	52	-
Ont.A	64	51	-	88
Ont.B	50	73	-	87
Va.A	70	76	36	83
Va.B	56	78	39	86
Mass.1st	80	63	40	-
" 2nd	90	70	40	-

cases.

PERFORMANCE AND I.Q.

Table VII shows the mean I.Q. of subjects with good patterns and poor patterns, for each group, for groups for which I.Q. scores were known. Seven out of twelve cases showed that best performance accompanies higher I.Q. We may then conclude that there is no marked tendency for students with higher I.Q. to perform better. Such a statement is not to be considered conclusive, due to the lack of thorough investigation of this subject.

D. VARIABILITY

By variability we mean variation in group response percentage to the same stimuli upon different administrations of the same test. While there are a number of statistically derived coefficients of variability, it seemed more suitable to the needs of this study to derive our own measure.

Two groups, Rochester A and Massachusetts, were each administered twice Tests IB, IIA, B, and IIIC. Administrations of the test for the Rochester A group were one week apart; for Massachusetts, considerably over a month. From this data (Appendices G, H, I, K), it was possible to tabulate for each stimulus, the percentage of change in group response percentages. (All percentages were based on the number of subjects in the group.) Our

TABLE VI

Performance and Sex

The Percentages of good patterns of male and female subjects for Tests IB, IIA.

Test IB

	Clarence				
	1st	2nd	Ont.A	Va.A	Va.B
% good, male	0.0	50	36	54	44
% good, female	100	80	81	82	64

Test IIA

	Clarence				
	1st	2nd	Ont.A	Va.A	Va.B
% good, male	17	50	36	69	67
% good, female	0.0	20	56	82	86

TABLE VII

Performance and IQ.

The Mean I.Q.[1] of subjects with good patterns, and with poor patterns, for Tests IB, IIA.

Test IB - Mean I.Q.

	Clarence					
	1st	2nd	Ont.A	Ont.B	Va.A	Va.B
good patterns	120	108	113	104	119	121
poor patterns	107	113	111	104	119	109

Test IIA - Mean I.Q.

	Clarence					
	1st	2nd	Ont.A	Ont.B	Va.A	Va.B
good patterns	101	115	111	106	120	118
poor patterns	111	106	112	97	117	110

[1] I.Q. scores cannot be compared between groups, since different groups used different tests. See Appendix D

measure of variability is thus the percentage of change in group response for each test item, based on data from two administrations of the test. Note that this is a measure of the variability of the group, and not of individuals. Thus for a given stimulus, x number of subjects might switch from response A to B, but if x other subjects reversed this trend, with a switch from B to A, there is no change in group response percentage, and variability is 0.0%.

Figure D gives the variability for Test IB for Mass. and Roch.A. Nine items show greater than 20% variability. Three of these can be attributed to the confusion of responses /a/ and /ar/ for Rochester, and thus indicate only that these responses are not separate categories for that group. The area of least reliability as shown in Fig. D is the area of the phoneme /ɔ/ (stimulus items 25, 22, 1, 19, 9, 20, 10; refer to Appendix B-3). Apart from these items, only two items (14, 18) were over 20%.

Figure E gives the variability for Test IIA for the two groups. Mass. shows considerably less variability than Roch.A. For Mass., five stimulus items show 15%, and all other items show 10% variability or less. The greater variability of Roch. is consistent with the poor performance of all groups from Western New York State (Table V), It is

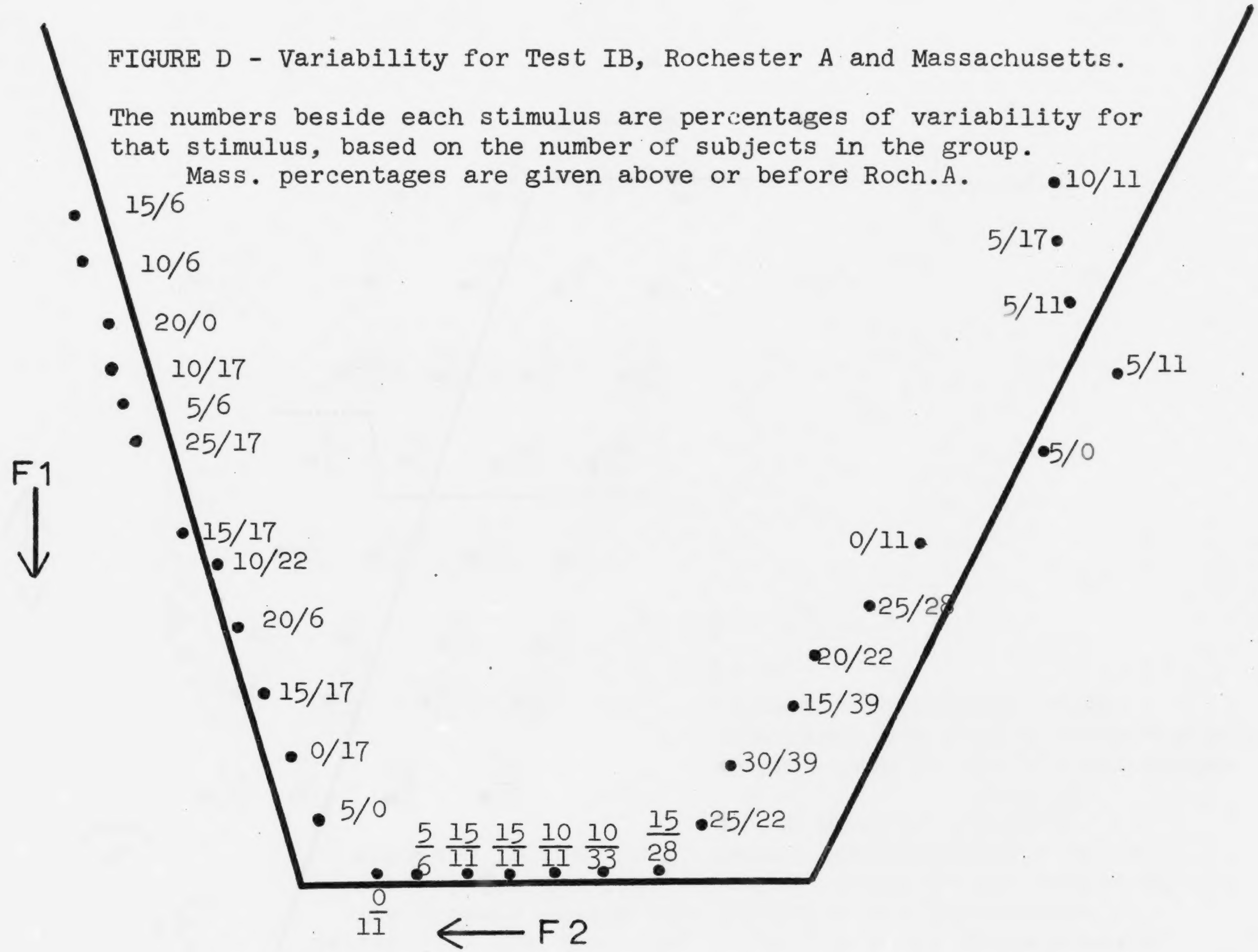
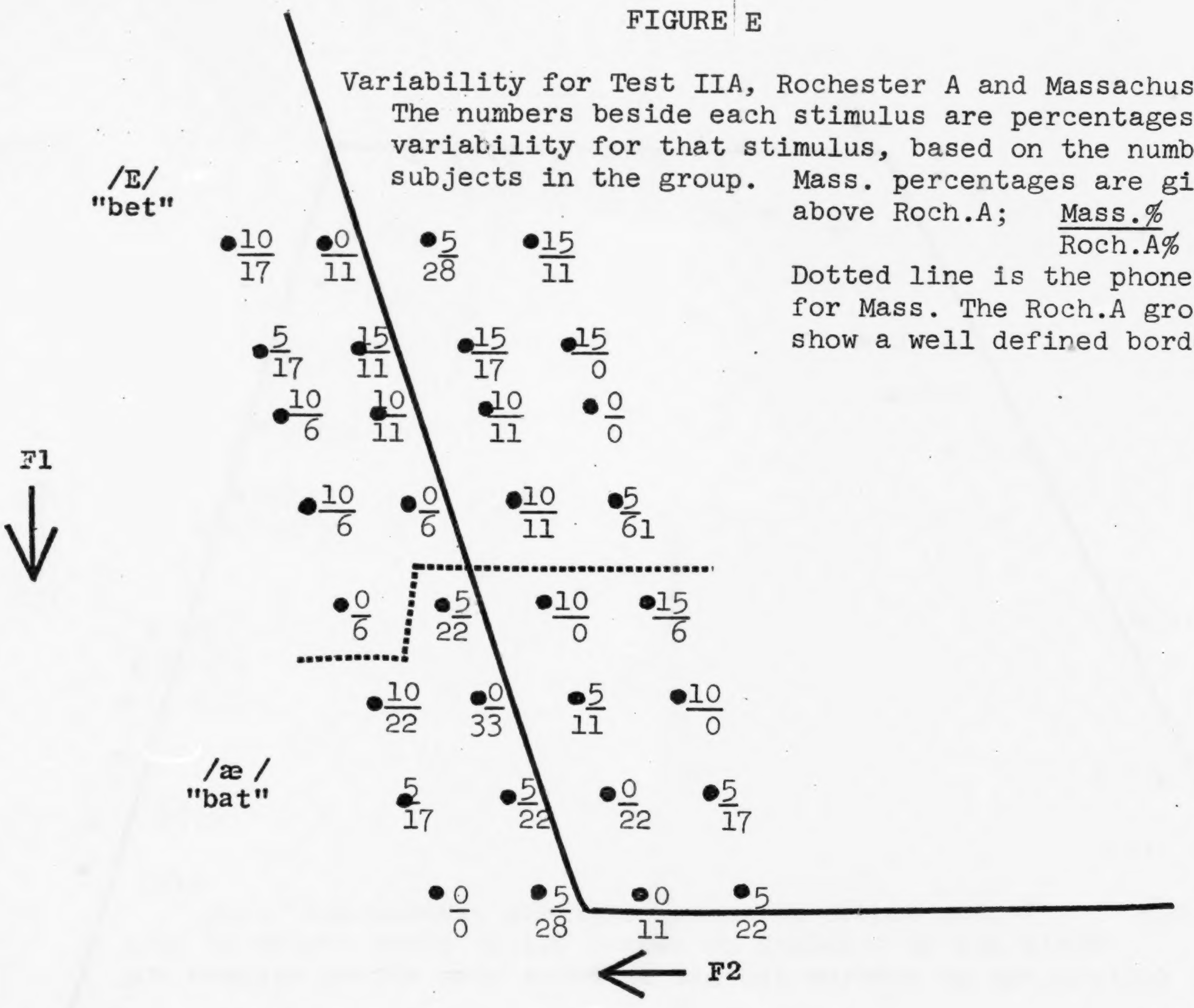


FIGURE D - Variability for Test IB, Rochester A and Massachusetts.

The numbers beside each stimulus are percentages of variability for that stimulus, based on the number of subjects in the group. Mass. percentages are given above or before Roch.A.

FIGURE E



Variability for Test IIA, Rochester A and Massachusetts. The numbers beside each stimulus are percentages of variability for that stimulus, based on the number of subjects in the group. Mass. percentages are given above Roch.A; $\frac{\text{Mass.\%}}{\text{Roch.A\%}}$

Dotted line is the phoneme border for Mass. The Roch.A group did not show a well defined border.

further evidence that they could not easily categorize "bet/bat."

Figure F gives variability for Test IIB for the two groups. Roch.A shows considerably less variability than Mass. Rochester percentages are, with five exceptions, 15% or less. The greater variability of Mass. is consistent with the fact that subjects in this group did not make the distinction "cot/caught" in their speech. (This is well established in dialect literature.[7] As additional evidence, Mass. subjects were asked after the test to write at the bottom of their answer sheets whether or not these words were pronounced differently. Virtually all replied with "no.")

Test IIC was not administered to Roch.A and Mass. Therefore, we do not have the above described measure of variability for this test. It is quite likely that Test IIC has a lower percentage of variability than Test IIA (15%), since performance on this test was extremely good (Table V) and phoneme boundaries (for all groups) are well defined. (cf. Appendix J)

We do have another sort of measure of variability for Test IIC. Nine subjects, (Roch.B) students at the University of Rochester, took the test twice, at an

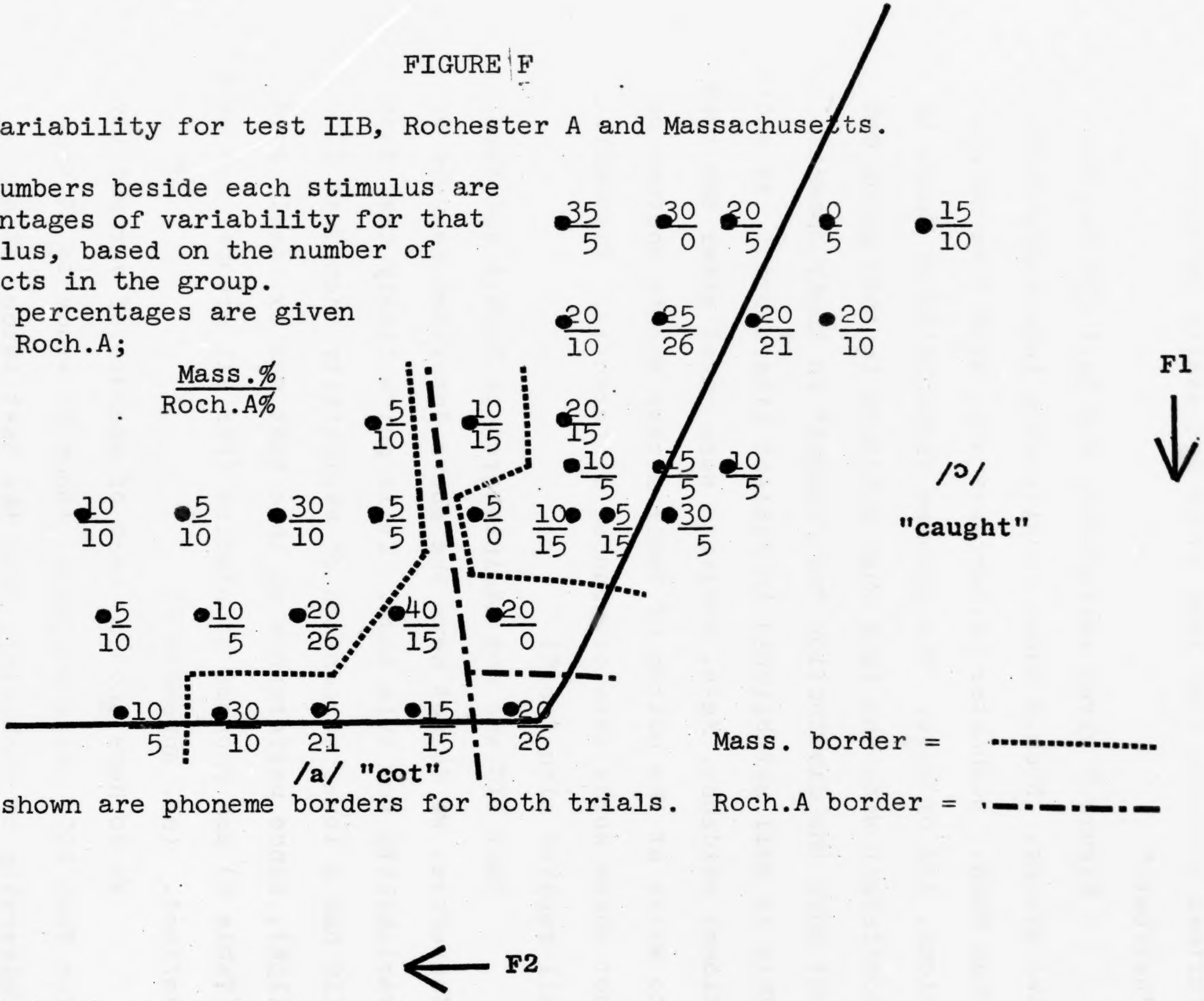
---

[7] See, for example, Kurath and McDavid, 1961.

FIGURE F

Variability for test IIB, Rochester A and Massachusetts.

The numbers beside each stimulus are percentages of variability for that stimulus, based on the number of subjects in the group.
Mass. percentages are given above Roch.A;



Also shown are phoneme borders for both trials.

interval of several days. Since the subjects were not tested as a group, we did not use the group variability measure described above. Instead we tabulated, for each stimulus item, the number of subjects who did not change response category, for two trials of the same test. Figure G gives the number of subjects, out of nine, who did not shift response. Note particularly on the lowest row of stimuli, the maximum number shifting response is 2/9 or 22%. It should be remembered that it is possible that such a shift produces no change in group response percentage. Thus the percentage of variability, as described for other tests, is probably well below 20% for the lowest row of stimuli for Test IIC.

Table VIII gives the variability for Test IIIC. It should be remembered that stimuli for Test IIIC are the same as for the lower row of Test IIC. Maximum average variability is 12%.

The above measures of variability are only approximate. We cannot say precisely to what extent the variability for one group tested twice will relate to a third testing of the same group. Nor is it clear how directly variability as measured for one group relates to another group. It has already been indicated that the reliability of a particular test for a group is probably related to the performance of the group in making that particular

FIGURE G

Individual Variability for Test IIC, for Rochester B, 9 Subjects

Shown here are the number of subjects, out of nine, with unchanged response for two trials.

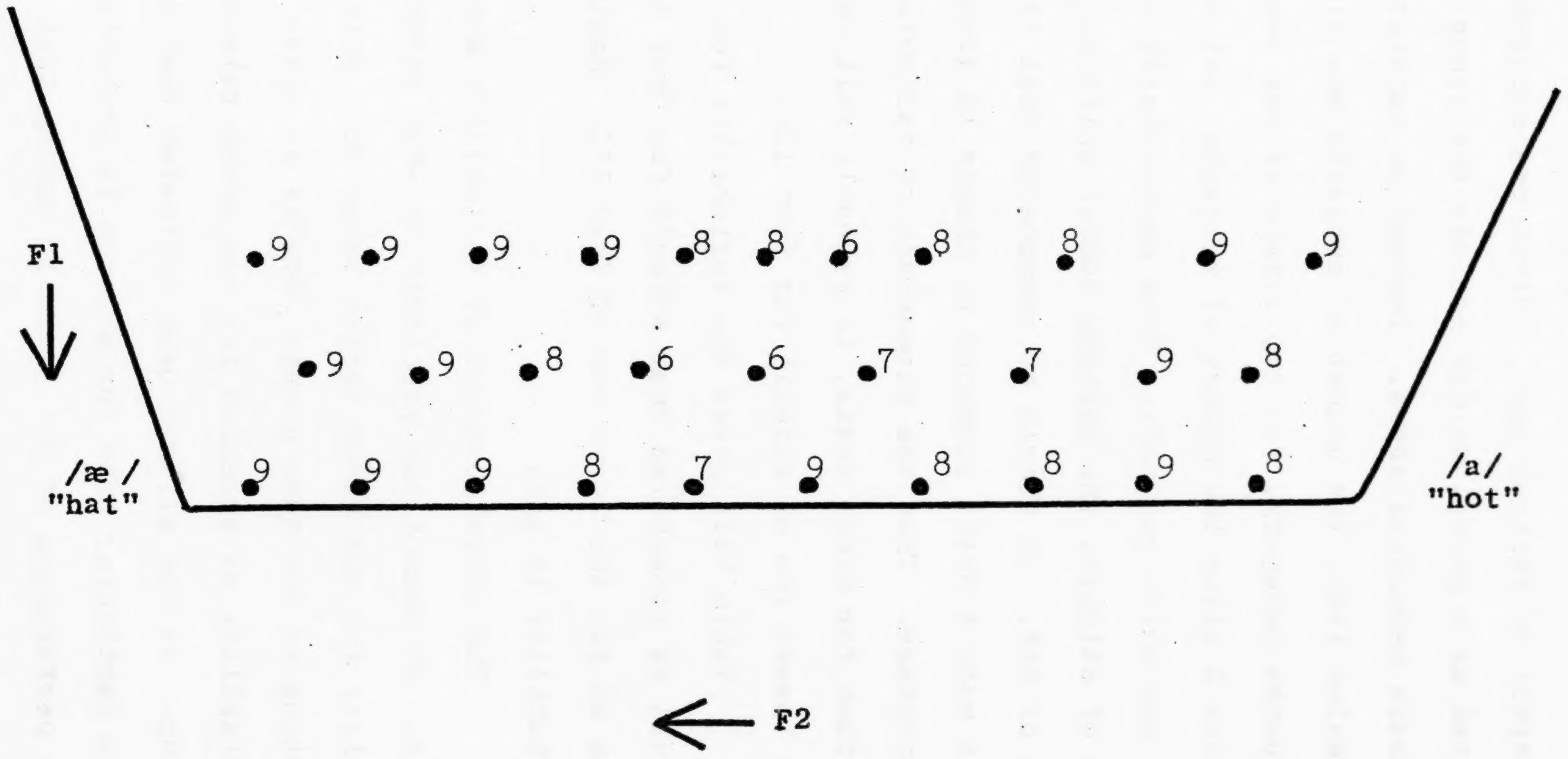


TABLE VIII

Variability for Test IIIC, Rochester A and Massachusetts.

Each column of numbers above a stimulus represents the percentages of variability for that stimulus, based on the number of subjects in the group. The lowest four numbers are the percentages for the four occurences of that stimulus; the uppermost number is the average percentage of variability for that stimulus.

Mass. percentages are given before Roch.A, Mass.%/Roch.A%.

Average %:	4/2	4/4	5/6	8/10	12/12	9/5	2/9	0/8
	0/0	5/5	5/5	5/10	10/15	5/5	0/0	0/0
	10/5	0/5	10/5	5/30	10/15	10/10	5/15	0/10
	5/0	5/5	5/15	15/5	5/10	15/0	5/10	0/10
	0/5	5/0	0/0	5/5	25/10	5/5	0/10	0/10
Stimulus Item	"hat" /ae/ a	b	c	d	e	f	g	"hot" /a/ h

F2

categorization. This is evidenced by the low reliability of Roch.A for Test IIA, and of Mass. for Test IIB.

Inspite of the approximate nature of this measure of variability, we shall refer to it later. For this reason we give below the following estimated percentages of variability, based on the analyses above. Since in every case it is the variability of the phoneme boundary that will concern us, we have disregarded high variability for stimuli not near a phoneme boundary. This is particularly relevant to Tests IIA, IIB.

For Test IB, the estimated variability is 20% or less, except in the area of the phoneme /ɔ/.

For Test IIA, the estimated variability is 15% or less, for groups with good performance (i.e. all groups but those from the Western New York region).

For Test IIB, the estimated variability is 20% or less, except for groups not having this contrast in their dialect.

For Test IIC, variability is estimated at 20% or less.

For Test IIIC, variability is estimated at 12% or less.

Estimates of variability, while based on data from Rochester A, Rochester B, and Massachusetts groups, will be considered to apply to all groups taking the tests.

E. INTER-GROUP DIFFERENCES

1. Concerning Methods of Analysis

In this section we shall analyze inter-group differences in selection of phoneme centers and differences in position of phoneme boundaries. We shall take note of all differences, without paying particular attention to those differences which might seem particularly of interest in relation to dialect. Only after we have made such an analysis, will we attempt to correlate these differences with previously established dialect features.

Differences in selection of phoneme centers are demonstrated by showing that some groups have phoneme centers which other groups do not have. This will be shown in analysis of the data from Test IB. A phoneme center has been defined (p. 2) as a point of maximum agreement in response category, i.e., a maximum percentage for a particular phoneme. We shall also use the term "phoneme area" to designate a number of stimuli heard as a single phoneme by a plurality of the subjects. Phoneme areas are enclosed by phoneme boundaries, as shown in Appendices F, G. According to these definitions, a phoneme area implies the existence of a phoneme center, and in most cases, the reverse is also true.

It must be remembered that Rochester A and Massachusetts took a forced choice version of Test IB, while other groups took an open-ended version. This may account for some of the differences between groups. Therefore, any comparison between these two groups, and other groups, must be done with this fact in mind.

Closely related to differences in selection of phoneme centers are differences in ability to categorize certain sets of stimuli. Thus if a group is unable to categorize /E/-/ae/, as required in Test IIA, we would not expect to find phoneme centers for both these phonemes, in the data from Test IB. Evidence for differences in ability to categorize in a given test is found: 1) in the study of performance on the test (cf. p. 40ff.), 2) in the variability for that test (cf. p. 48ff.), and 3) in the examination of phoneme boundaries, to determine whether they are well defined. (A well defined boundary is one in which the shift in percentages of response across the boundary is an abrupt shift.)

In analyzing differences in position of phoneme boundaries, our procedure is first, to observe the positions of these boundaries relative to each other, as shown in the Appendices F-K. Secondly, we make a judgement as to the reliability of these differences, based upon the variability of the particular test. In analyzing Test IB, we shall select a particular stimulus item, for which intergroup differences in response illustrate particularly well the differences in position of a phoneme boundary. This procedure allows the use of the chi-square test of significance of the inter-group differences.[7]

2. Differences in Selection of Phoneme Centers (Test IB)

An examination of Appendix G shows that all groups have well defined phoneme centers for /i/,/ae/,/o/, and /u/. Choice among the phonemes /e/,/E/,/a/, and /ɔ/, varies considerably by group. Table IX shows the differences in selection of phoneme centers (/e/,/E/,/a/,/ɔ/), for eight groups taking Test IB.

Ontario groups favor /e/, as do Buffalo groups to a lesser degree. (The /e/ percentages for Buffalo are considerably lower than those for Ontario.) Both Virginia and Massachusetts favor /e/, but one Virginia group, and one trial of Massachusetts, also show an area of split, /E/e/.[8] Rochester shows the split /E/e/ in both trials.

---

[7] The judgments of significance and variability made for these tests are at least sufficient for the needs of this study. By taking as our percentage of variability the higher of the observed percentages, we increase the probability that the groups studied would show less than this variability, if tested a second time. We did not, however, test each group twice.

The mathematical model from which chi-square was derived is not entirely satisfied by the data of these tests. Primarily this means that our subjects do not represent a random sample of a larger population. In spite of this limitation on the interpretation of our results, we are justified in using the test as a test of significant difference.

[8] The category /E/e/ refers to a phoneme area where,

TABLE IX

Differences in Selection of Phoneme Centers for Eight Groups Taking Test IB.

(Open-ended Version of Test IB)

	/e/	/E/	/E/e/	/a/	/ɔ/
Buf.A	35	-	-	-	-
Buf.B	50	-	-	-	75
Ont.A	82	39	-	-	86
Ont.B	75	-	-	-	70
Va.A	-	43	36/50	53	63
Va.B	-	65	-	52	65

(Forced Choice Version of Test IB)

				/ar/		
Roch.A						
1st Trial	-	-	44/39	-	56	-
2nd Trial	-	-	28/56	-	83	-
Mass.						
1st Trial	-	75	65/30	90	75	55
2nd Trial	-	60	-	100	90	40

Explanation:

Shown here are the highest percentages of agreement for the phoneme areas /e/,/E/,/a/,/ɔ/. A dash indicates that there was no area for that phoneme.

The category /E/e/ indicates a phoneme area where, for each stimulus, responses were approximately evenly split between the two categories.

The category /ar/ refers to the response "tar," and was tabulated only for the forced choice version of Test IB.

Massachusetts shows a center for /ar/ (="tar"), which is not found in the Rochester A data. The choice /ar/ was analyzed only for the forced choice version of Test IB, taken only by Rochester A and Massachusetts.

Only Virginia and Massachusetts show high percentages for both the /a/ and /ɔ/ phoneme centers. Only Virginia has both phoneme areas well established. By this we mean that, referring to Appendix G, Virginia groups have at least three stimuli in the area of each phoneme, /a/ and /ɔ/, while Massachusetts groups both have only one stimulus in the phoneme area of /ɔ/.

3. Inability to Categorize Certain Sets of Stimuli (Tests IIA, IIB)

Only Rochester A and Buffalo groups have difficulty categorizing the stimuli of Test IIA. Proof of this lies in: 1) the low performance of these groups (Table V), 2) the higher variability of Rochester A as compared to Massachusetts (Figure E), and lastly 3) the lack of well defined phoneme boundaries for Rochester A and Buffalo groups (Appendix H).[9] (These boundaries are considered less well

---

for each stimulus, responses were approximately evenly split between the two categories.

[9] The failure of Rochester A and Buffalo groups to categorize /E/-/ae/ in Test IIA is consistent with the lack of phoneme centers for /E/ for these groups in Test IB. There is also some indication of a split /E/ae/ for Rochester, Test IB. In general, however, the inability to categorize /E/-/ae/ is not so clearly evidenced in Test IB as it is in Test IIA.

defined than those of other groups because there are very few percentages over 80%, and very few under 25%.)

Test IIB gives evidence for the inability to categorize /a/-/ɔ/, but the inter-group differences are less striking than for Test IIA. Performance on this test for all groups was poor. (Table V) As concerns the patterns shown in Appendix I, the Virginia groups stand out from the rest, in that they are the only groups with well defined phoneme boundaries between /a/ and /ɔ/. This is consistent with evidence from Test IB, that Virginia was best able to distinguish these phonemes.

It will be remembered that in Test IB, Massachusetts had both an /a/ and /ɔ/ center, whereas Rochester lacked the /a/ center. This might be given as evidence that the Rochester group did not make the /a/-/ɔ/ categorization. Inconsistent with this interpretation, the variability for Rochester A, Test IIB (Figure F), was lower than Massachusetts. Further evidence that Massachusetts had difficulty with the /a/-/ɔ/ categorization is the fact that the Massachusetts group shows an extreme shift in phoneme boundary for the two trials of Test IIB, compared to the relatively stable Rochester A boundary. Seen in this light, it is possible that the phoneme boundaries for the Massachusetts subjects are artifacts of the testing situation, and that while they are capable of making the categorization /a/-/ɔ/, these subjects show no fixed boundary which characterizes them as a group.

4. Differences in Position of Phoneme Boundaries

There were noticeable differences in position of phoneme boundaries for the phoneme pairs /i/-/e/,/E/-/ae/, /ae/-/a/,/a/-/ɔ/, and /o/-/u/.

The Phoneme Boundary /i/-/e/.

Differences in position of the /i/-/e/ boundary are most apparent if we consider responses to stimulus number 29, Test IB. (Appendix B-3 gives the position of this stimulus on the F1/F2 chart.) Table X(a) shows the number of responses /i/ and /I/, combined, for stimulus number 29, Test IB, for eight groups.[10] Table X(b) shows which inter-group differences from Table X(a) are significant according to chi-square. Table XI gives the same information for /e/ responses to this stimulus. The inter-group differences shown in these two tables may be summarized as follows:

a. No groups differ significantly from another group from the same immediate area, (e.g., Buf.A and Buf.B), nor do first and second trials of the same group differ

---

[10] It seemed justifiable to include the few /I/ responses in with the /i/ responses, since in no case did a group show a distinct phoneme area for /I/. The /I/ responses were found distributed relatively evenly in with the /i/ responses. We assume that the primary distinction between these two phonemes is not one of formant frequency, but rather of duration.

TABLE X(a)

The Number of /i/ and /I/ Responses to Stimulus Number 29, Test IB, for Eight Groups.

	Buf.A	Buf.B	Roch.A		Ont.A	Ont.B	Va.A	Va.B	Mass.	
			1	2					1	2
/i/,/I/	11	5	13	13	4	2	15	17	8	11
other	9	11	5	5	24	22	15	6	12	9

TABLE X(b)

Inter-group Differences in /i/ and /I/ Responses to Stimulus number 29, Test IB, for the Eight Groups.

X indicates an inter-group difference significant at p = .05, according to chi-square.

	Buf.A	Buf.B	Roch.A		Ont.A	Ont.B	Va.A	Va.B	Mass.	
			1	2					1	2
Buf.A					X	X				
Buf.B			X	X				X		
Roch.A										
1					X	X				
2					X	X				
Ont.A							X	X		X
Ont.B							X	X	X	X
Va.A										
Va.B										
Mass.										
1										
2										

TABLE XI(a)

The Number of /e/ Responses to Stimulus Number 29, Test IB, for Eight Groups

	Buf.A	Buf.B	Roch.A 1	Roch.A 2	Ont.A	Ont.B	Va.A	Va.B	Mass. 1	Mass. 2
/e/	5	4	1	1	23	16	8	1	5	4
other	15	12	17	17	5	8	22	22	15	16

TABLE XI(b)

Inter-group Differences in /e/ Responses to Stimulus Number 29, Test IB, for Eight Groups.

X indicates an inter-group difference significant at p = .05, according to chi-square.

	Buf.A	Buf.B	Roch.A 1	Roch.A 2	Ont.A	Ont.B	Va.A	Va.B	Mass. 1	Mass. 2
Buf.A					X	X				
Buf.B					X	X				
Roch.A 1					X	X				
Roch.A 2					X	X				
Ont.A							X	X	X	X
Ont.B							X	X	X	X
Va.A										
Va.B										
Mass. 1										
Mass. 2										

significantly.

b. Ontario groups contrast with all groups in preference for /e/. (cf. Table XI) Ontario also contrasts with all groups in number of /i/ and /I/ responses, (Table X), with the exception of Buffalo B and Massachusetts, First Trial. Buffalo B shows a low number of /i/ and /I/ responses, contrasting with Rochester A and Virginia B.

The relatively low percentages of /i/ and /I/ responses for Buffalo B do not indicate for this group a switch to /e/ or any other phoneme, but only reflect poor performance on this stimulus for this group. The highest percentage for any phoneme, for Buffalo B, for stimulus number 29, is 25%. (cf. Appendix G-2)

The reliability of the contrast between Ontario and other groups is determined by reference to the study of variability. (p. 58) We take as the maximum probable shift of percentage for Test IB, 20%. It is thus highly probable that if any two percentages differ by 40% or more, then the greater of the two percentages will remain greater upon repeated testing. By this criterion of 40%, the Ontario preference for /e/, (82% and 66%), is reliable. Rochester A preference for /i/ (72%), is also reliable. Massachusetts preference for /i/ (40%, 55%), is questionable, except for the fact that this group was tested twice, with the same preference both times. Virginia B preference for /i/ (56%), is nearly 40% greater than any other percentage for that stimulus. (The second highest preference is for /I/, 17%) Buffalo and Virginia A preferences for /i/ are unreliable. (40% and 36% respectively)

The Phoneme Boundary /E/-/ae/

Figure H shows for Test IB the position of the /E/-/ae/ boundary for eight groups, based on data in Appendix G. Differences in the /E/-/ae/ boundary are most apparent if we consider responses to stimulus number 18, Test IB. (Appendix B-3 gives the position of this stimulus on the F1/F2 chart) Rochester A and Buffalo have the highest boundaries, and Ontario and Virginia the lowest. These differences, however, are based on very slight differences in percentages, and their significance and reliability must be questionned. Table XII(a) shows the number of /ae/ responses for stimulus 18, Test IB, for eight groups. Table XII(b) shows which inter-group differences from Table XII(a) are significant according to chi-square. We would expect Rochester A and Buffalo groups to prefer /ae/, in contrast to groups with lower boundaries. The inter-group differences shown in Table XII may be summarized as follows:

a. No groups differ significantly from other groups from the same immediate area, nor do first and second

FIGURE H

The Position of the /E/-/ae/ Phoneme Boundary for Eight Groups Taking Test IB. Based on Appendix G.

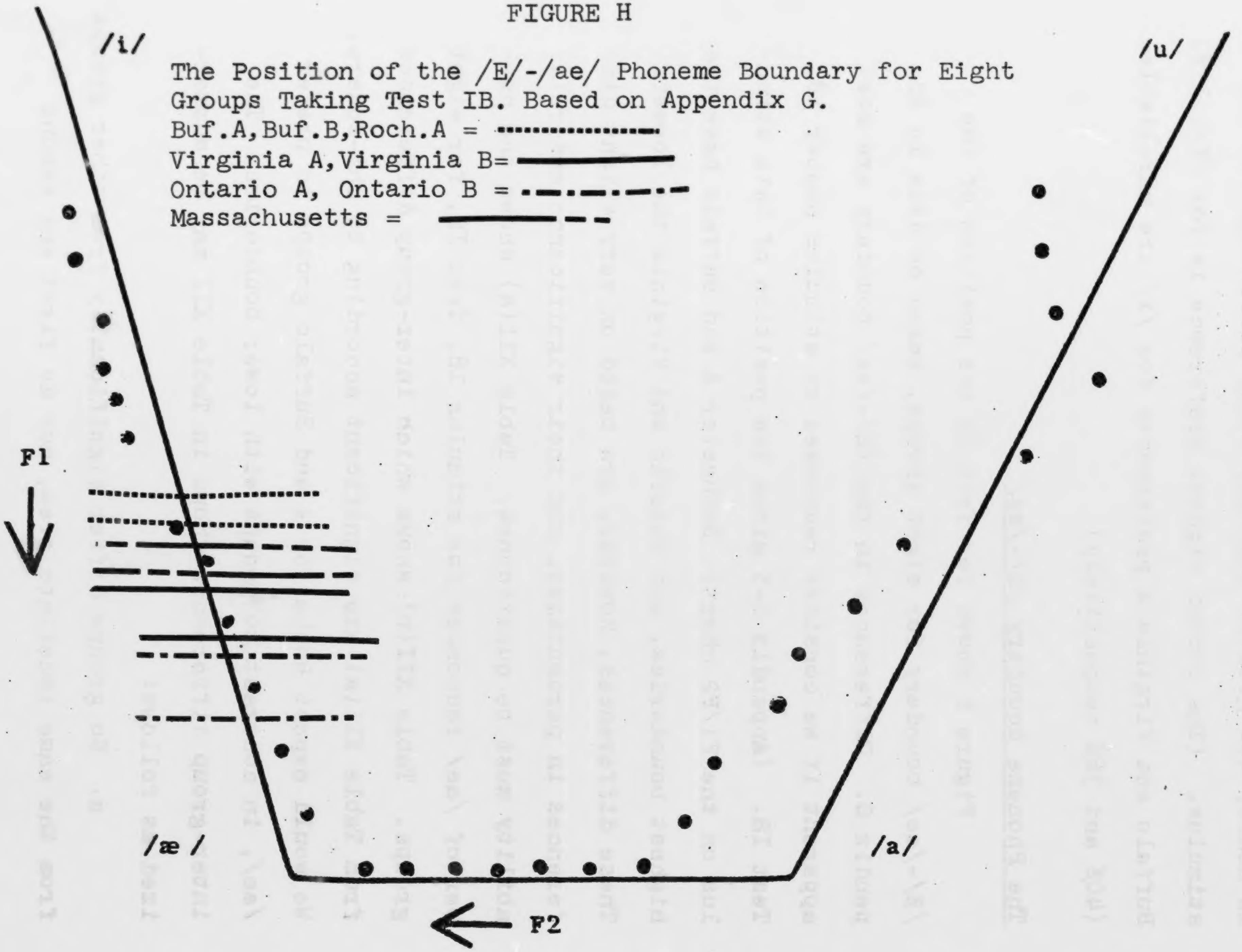


TABLE XII(a)

The Number of /ae/ Responses to Stimulus Number 18, Test IB, for Eight Groups.

	Buf.A	Buf.B	Roch.A		Ont.A	Ont.B	Va.A	Va.B	Mass.	
			1	2					1	2
/ae/	9	7	6	10	3	5	5	1	8	10
other	11	9	12	8	25	19	25	22	12	10

TABLE XII(b)

Inter-group Differences in /ae/ Responses to Stimulus Number 18, Test IB, for Eight Groups.

X indicates an inter-group difference significant at p = 0.5, according to chi-square

	Buf.A	Buf.B	Roch.A		Ont.A	Ont.B	Va.A	Va.B	Mass.	
			1	2					1	2
Buf.A					X			X		
Buf.B					X			X		
Roch.A										
1								X		
2					X	X	X	X		
Ont.A									X	X
Ont.B										
Va.A										X
Va.B									X	X
Mass.										
1										
2										

trials of the same group differ significantly. Similarly, groups with highest boundaries (Buffalo and Rochester A) do not differ significantly among each other, nor do groups with lowest boundaries. (Ontario and Virginia)

b. The inter-group differences between Rochester A and Buffalo on the one hand, and Virginia and Ontario on the other, predictable from Figure G, are not generally significant. The most striking contrasts are between Rochester A and Virginia, and between Virginia and Massachusetts. For both pairs of groups, three out of four combinations contrast significantly.

We conclude that the inter-group differences in position of the /E/-/ae/ boundary, for Test IB, which are observable in Figure H are of doubtful statistical significance.

The reliability of the preferences for /ae/ is determined in the same manner as for /i/-/e/ above. Using the criterion of 40%, none of the preferences for any phoneme by any group, for stimulus 18, is reliable, with the unique exception of the Virginia B preference for a category other than /ae/. (5% /ae/, 47% /E/)

Figure I shows the positions of the /E/-/ae/ boundaries for the eight groups taking Test IIA. The boundaries for Rochester A and Buffalo tend to drop obliquely towards the lower right, dividing the stimuli into backed stimuli

(/E/), and fronted stimuli (/ae/). (Figure I-1)[11] Massachusetts, Ontario, and Virginia borders run horizontally and drop to the lower left, dividing the stimuli into raised stimuli (/E/), and lowered ones (/ae/), along nearly congruent lines. (Figure I-2) The slight differences between the Virginia, Massachusetts, and Ontario boundaries are not reliable.

The Phoneme Boundary /ae/-/a/

There are no apparent differences in /ae/-/a/ boundary in Test IB. This is at least in part because percentages were extremely low in this area of the F1/F2 chart, for Test IB, and because percentages for /a/ were in general very low for all groups.

Figure J shows the position of the /ae/-/a/ boundary for the six groups taking Test IIC. Buffalo contrasts with Ontario in being more to the left, with Virginia lying between the two.

The reliability of the contrast between Ontario and Buffalo is determined by reference to the study of variability. (p. 58) Since the maximum probable variability for Test IIC is 20%, we take as our criterion of reliability a difference of 40%. Table XIII gives the percentages of

---

[11] Although the Rochester A group showed high variability for this test, the tendency toward a division into fronted and backed sounds appears in both trials. For this reason we consider this tendency to be reliable.

FIGURE I-1

Position of the /E/-/ae/ Phoneme Boundary for Eight Groups Taking Test IIA. Based on Appendix H.

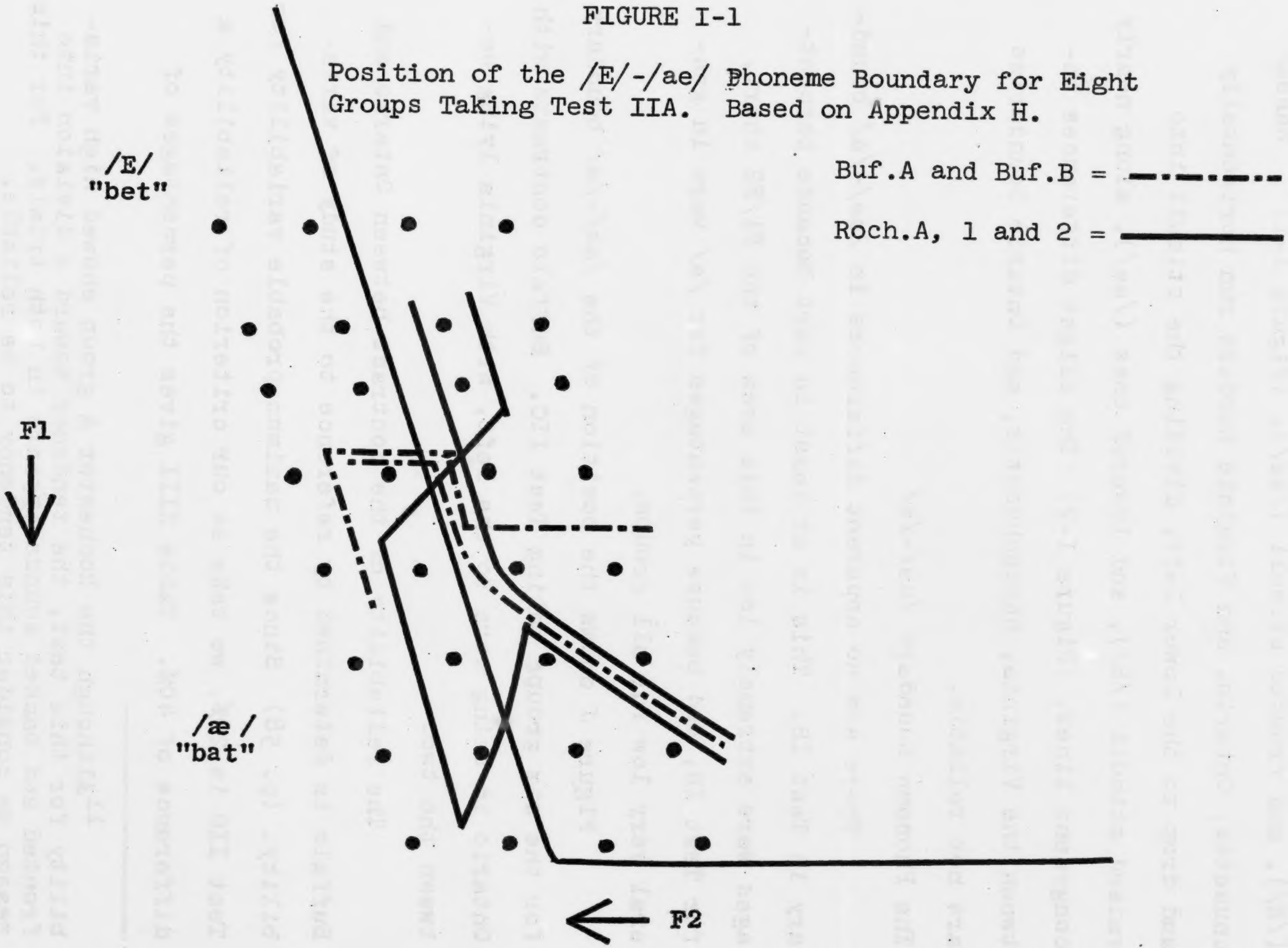


Figure I-2

Position of the /E/-/ae/ Phoneme Boundary for Eight Groups Taking Test IIA. Based on Appendix H.

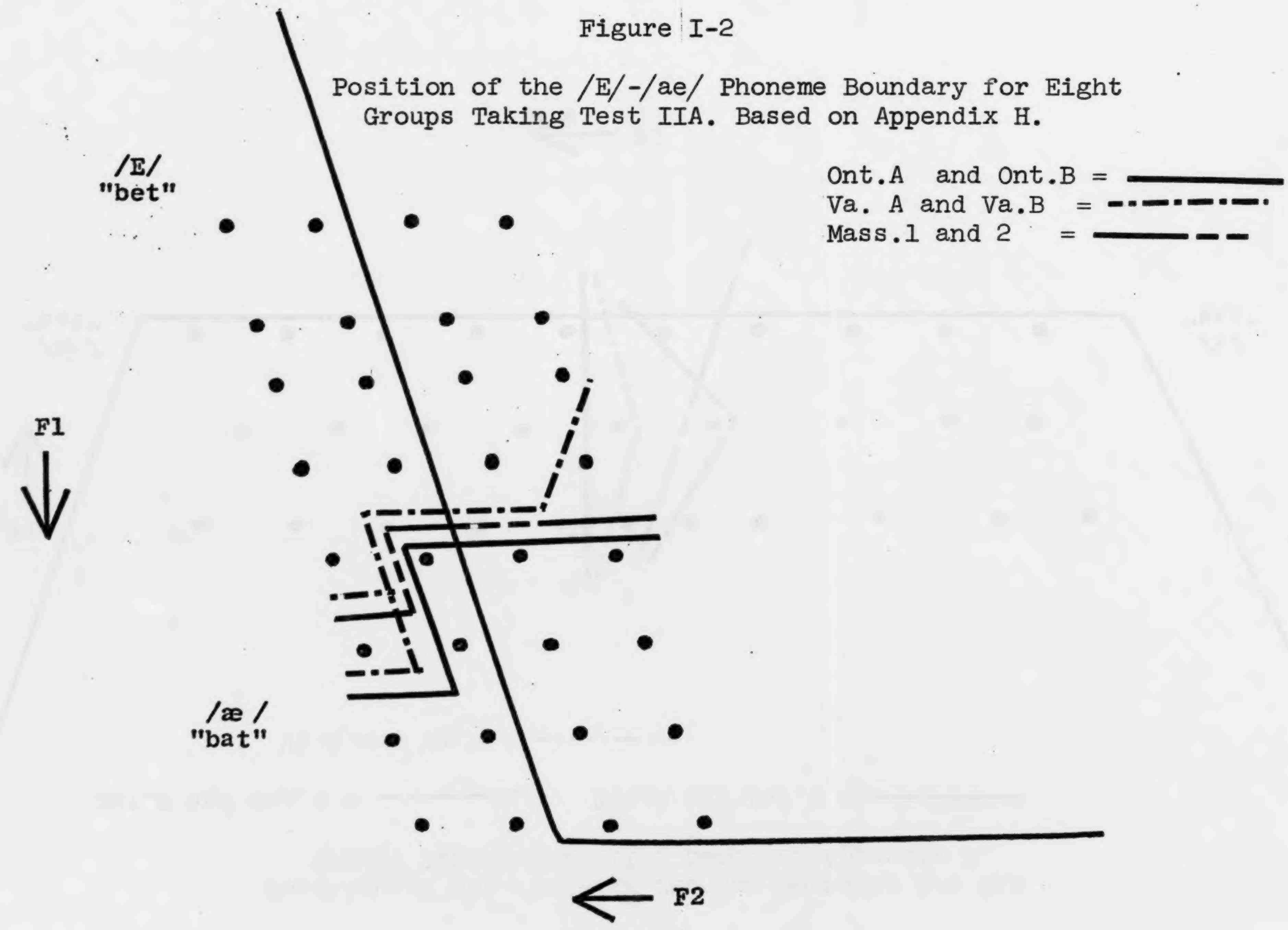
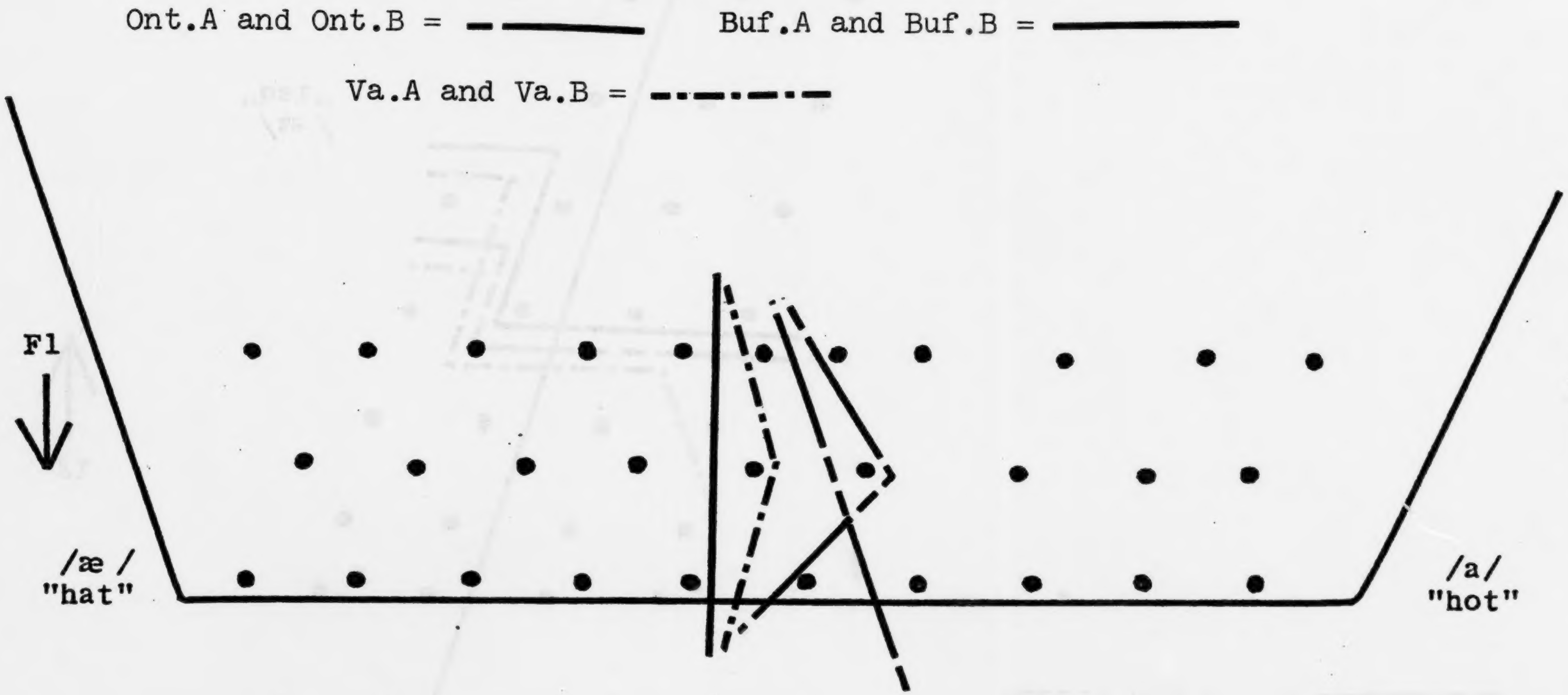


FIGURE J

Position of the /ae/-/a/ Phoneme Boundary for Six Groups Taking Test IIC. Based on Appendix J.



/ae/ response for the six stimuli that determine the phoneme boundaries for Buffalo and Ontario. If we compare the highest Buffalo percentages with the lowest Ontario percentages, then none of the stimuli show reliable differences in response, (although in every case Buffalo percentages are below those of Ontario.). Stimulus 12 comes nearest to a reliable contrast between the two groups.

TABLE XIII

Percentages of /ae/ Response, for Buffalo and Ontario, Test IIC, for the Six Stimuli which Determine the Phoneme Boundary.

	Stimulus Number					
	11	18	7	10	12	27
Buf.A	61	76	23	47	8	23
Buf.B	75	50	25	18	6	12
Ont.A	85	96	40	74	59	41
Ont.B	96	100	37	79	42	54

Note: Appendix B-6 gives the position of these stimuli on the F1/F2 chart.

Since the differences in boundaries were so slight, it is of interest to consider whether responses to stimulus 12 (which shows the most reliable contrast) show a significant difference according to the chi-square test. Table XIV shows the number of /ae/ responses to stimulus 12, Test IIC, for Ontario and Buffalo groups. Test by chi-square shows the contrast between Ontario groups and Buffalo groups in every case to be significant. Differences between groups in the same region are not significant.

TABLE XIV

Number of /ae/ Responses to Stimulus 12, Test IIC, for Buffalo and Ontario groups.

	Buf.A	Buf.B	Ont.A	Ont.B
/ae/	2	1	16	10
other	19	15	11	14

We conclude that the differences in position of the /ae/-/a/ boundary, between Buffalo and Ontario, do not entirely withstand the test of reliability. Nevertheless the differences are significant and it seems clear that there is a tendency for Buffalo groups to place the /ae/-/a/ boundary more to the left than do Ontario groups. As a further indication of this tendency, we note that Buffalo groups show a total number of /ae/ responses less than 50% (48% and 41%), while Ontario groups show a total number of /ae/ responses well over 50% (55% and 56%).[12] This preference for /ae/ by Ontario is consistent with a more backed

---

[12] We have no measurement of variability for the total number of responses for Test IIC. For Test IIA, and for Test IIB, variability in two trials for Rochester A and Massachusetts was never more than 2%. Thus, we suppose that the lower percentages of /ae/ shown by Buffalo groups are a reliable measure.

/ae/-/a/ border for these groups.

Test IIIC shows quite clearly that there is a reliable inter-group difference in position of the /ae/-/a/ boundary, when comparing Rochester A and Massachusetts. Figure K shows the positions of the /ae/-/a/ boundary, for two trials each of Rochester A and Massachusetts groups. The position of the boundaries between stimuli is determined by the intersection of the /ae/ and /a/ curves, (Appendix K) and is an interpolation of the percentages for the stimuli on either side of the boundary. Thus the positions of boundaries for Test IIIC are much more precisely determined than for, say Test IIC, where the exact position of a border between stimuli is not determined. Moreover, the tendency for Rochester A to show a boundary to the left of Massachusetts continues to appear upon repeated testing.

The Phoneme Boundary /a/-/ɔ/

Test IB gives little indication of differences in position of the /a/-/ɔ/ phoneme boundary, since only Virginia groups show well defined areas for both these phonemes.

Figure L shows the positions of the /a/-/ɔ/ boundary for the seven groups taking Test IIB. Buffalo contrasts with Virginia, and Rochester A lies between these two. Massachusetts shows an extreme shift in position of the

FIGURE K

Position of the /ae/-/a/ Phoneme Boundary for Two Groups Taking Test IIIC. Based on Appendix K.

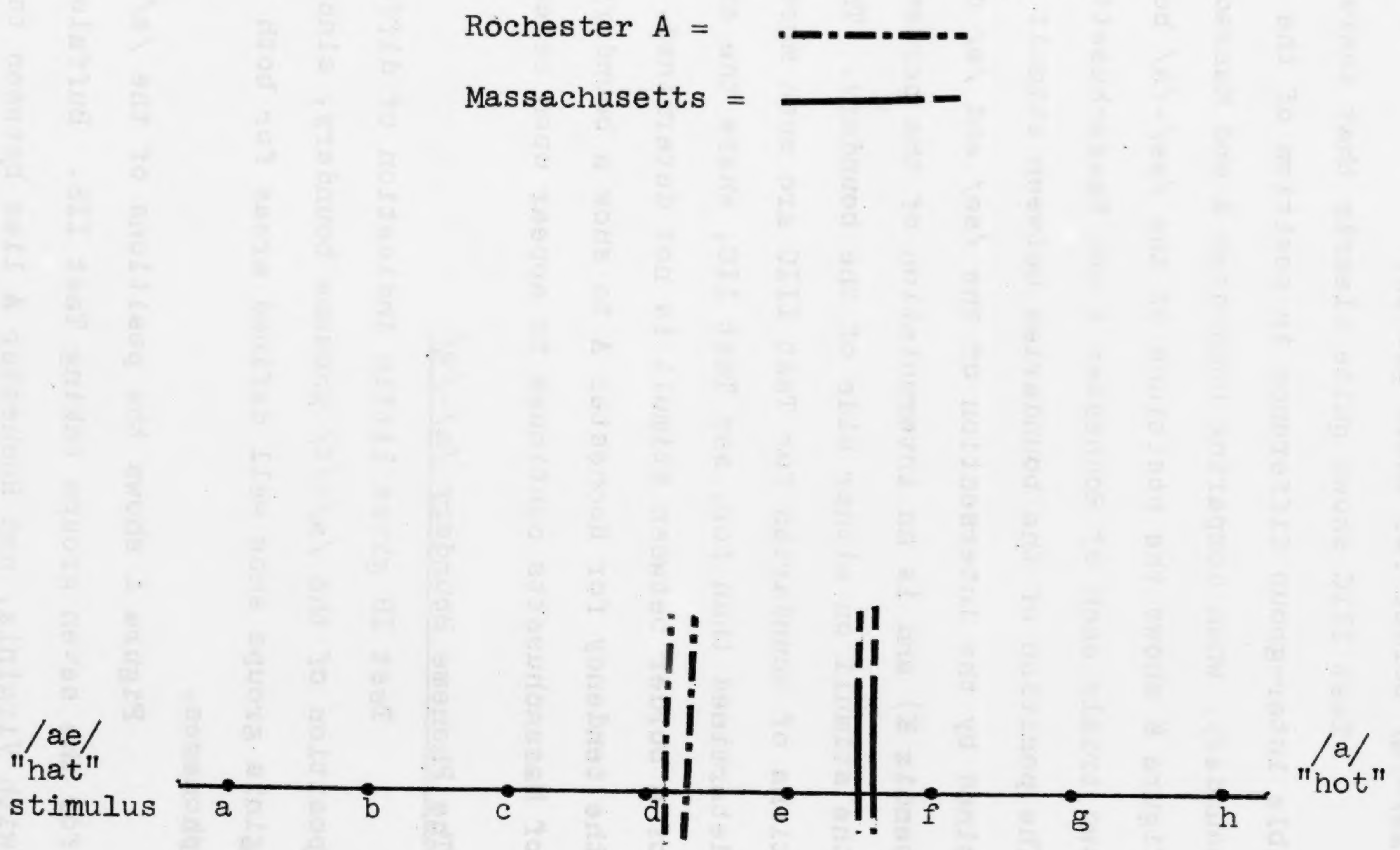
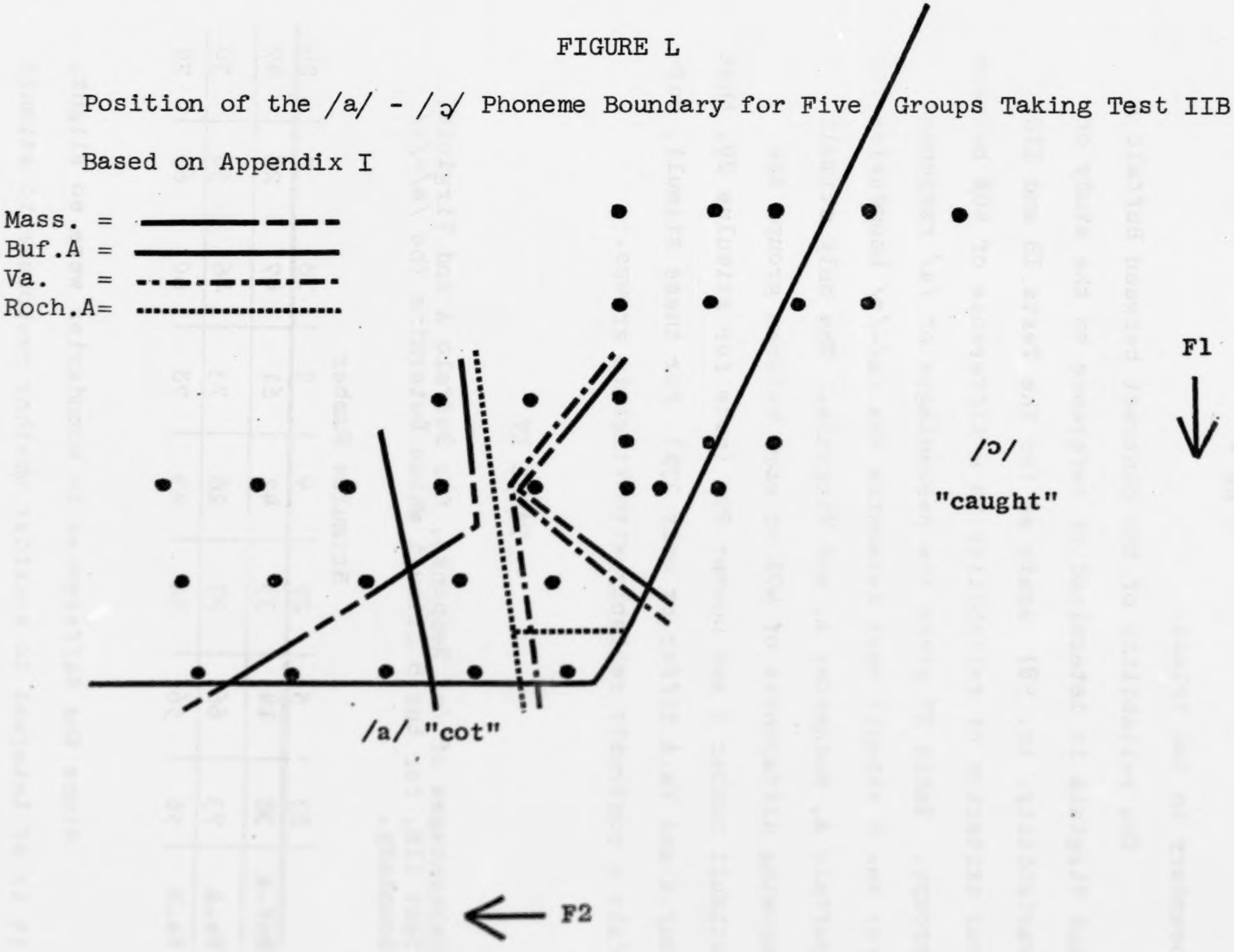

Rochester A =
Massachusetts =
/ae/
"hat"
stimulus
a
b
c
d
e
f
g
h
/a/
"hot"

FIGURE L

Position of the /a/ - /ɔ/ Phoneme Boundary for Five Groups Taking Test IIB.

Based on Appendix I



boundary in two trials.

The reliability of the contrast between Buffalo A and Virginia is determined by reference to the study of variability. (p. 58) Again as for the Tests IB and IIC, our criterion of reliability is a difference of 40% between groups. Table XV gives the percentages of /a/ responses for the 8 stimuli that determine the /a/-/ɔ/ boundaries for Buffalo A, Rochester A, and Virginia. The only stimuli showing differences of 40% or more between groups are stimuli number 5 and number 29. (note for stimulus 29, that Buf.A and Va.A differ by only 37%) For these stimuli, Buffalo A contrasts reliably with Virginia groups.

TABLE XV

Percentages of /a/ Response, for Buffalo A and Virginia, Test IIB, for the 8 Stimuli which Determine the /a/-/ɔ/ Boundary.

	Stimulus Number							
	23	5	29	9	2	18	8	24
Buf.A	38	14	33	42	61	47	28	47
Va.A	73	66	70	26	73	46	56	30
Va.B	56	56	86	43	78	60	69	74

Since the differences in boundaries were so slight, it is of interest to consider whether responses to stimuli 5 and 29 (which show reliable contrasts) show a significant difference according to the chi-square test. Table XVI shows the number of /a/ responses to these stimuli, for Buffalo A and Virginia groups. Test by chi-square shows the contrast between Buffalo A and Virginia groups in both cases significant at $p = .05$. Differences between Virginia A and Virginia B are not significant.

TABLE XVI(a)

Number of /a/ Responses to Stimulus 5, Test IIB, for Buffalo A and Virginia groups.

	Buf.A	Va.A	Va.B
/a/	3	20	13
other	18	10	10

TABLE XVI(b)

Number of /a/ Responses to Stimulus 29, Test IIB, for Buffalo A and Virginia groups.

	Buf.A	Va.A	Va.B
/a/	7	21	20
other	14	9	3

The Phoneme Boundary /o/-/u/

Differences in position of the /o/-/u/ boundary are most apparent if we consider responses to stimulus number 28, Test IB. (Appendix B-3 shows the position of this stimulus on the F1/F2 chart.) Table XVII shows the number

TABLE XVII(a)

The Number of /u/ Responses to Stimulus Number 28, Test IB, for Eight Groups

	Buf.A	Buf.B	Roch.A		Ont.A	Ont.B	Va.A	Va.B	Mass.	
			1	2					1	2
/u/	15	7	13	16	6	6	20	15	9	10
other	5	9	5	2	22	18	10	8	11	10

TABLE XVII(b)

Inter-group Differences in /u/ Responses to Stimulus Number 28, Test IB, for Eight Groups

X indicates an inter-group difference significant at p = 0.5, according to chi-square

	Buf.A	Buf.B	Roch.A		Ont.A	Ont.B	Va.A	Va.B	Mass.	
			1	2					1	2
Buf.A					X	X				
Buf.B				X						
Roch.A										
1					X	X				
2					X	X			X	X
Ont.A							X	X		
Ont.B							X	X		
Va.A										
Va.B										
Mass.										
1										
2										

number of responses /u/, and the inter-group differences significant according to chi-square, for stimulus number 28, Test IB, for eight groups. The inter-group differences shown in these two tables may be summarized as follows:

a. No groups differ significantly from another group from the same immediate area, nor do first and second trials of the same group differ significantly.

b. Ontario groups contrast with Rochester A and Virginia groups. Ontario groups favor /o/, whereas Rochester A and Virginia favor /u/.

c. Buffalo B differs significantly from Rochester A, second trial, and Rochester A, second trial differs from Massachusetts. This contrast is not because Buffalo B and Massachusetts show any particularly strong preference for /o/, but because these two groups split about evenly between /u/ and /o/, whereas Rochester A, first trial shows an unuaually strong preference for /u/.

F. EXPLANATION OF INTER-GROUP DIFFERENCES BY REFERENCE TO DIALECTAL FEATURES.

1. Previously Established Features of the Dialects of Western New York, Massachusetts, Ontario, and Virginia.

The following are dialectal features of the pronunciation of vowels, for the four dialects studied.

a. /e/,/o/: The dialects of Canada in general show certain Scottish characteristics. (Ahrend, 1934, p. 136)

One feature of Scottish dialect is a tendency not to diphthongize /e/ and /o/. (Wise, 1957, p. 258) These facts argue in support of our own observation that the vowels of "eat" and "two" are less diphthongized in Ontario speech, than in most American dialects. There is, however, no authoritative statement to this effect, in the literature.

b. /E/: There is a tendency in some American dialects, particularly in the South, to diphthongize and raise /E/, so as to resemble /e/, though the two sounds generally remain distinct and in contrast in pronunciation. (Kurath and McDavid, 1961, p. 132ff.)

c. /ae/: The raised /ae/ is commonly attributed to the dialects of Buffalo and Rochester. This raising is noted in Kurath and McDavid, for the word "bag." Natives of the region often mention this raising of /ae/ to resemble /E/ as a feature of their own dialect. However, the raising of /ae/ is also a socially determined feature.[13] (Thomas, 1947, p. 69; Bronstein, 1960, p. 156)

d. /a/: The phoneme /a/ has variants which are extremely fronted for Buffalo and Rochester, and extremely backed for Virginia.[14]

---

[13] During the time of this study, we ran a short experiment with five subjects from Ft. Erie, Canada, and five subjects from Rochester, N.Y., to determine the relative degree of raising of /ae/ and fronting of /a/ for both groups. Subjects were asked to read a passage containing words with phonemes /ae/ and /a/. These words were spliced out and made into a listening test, in which five trained

e. /ar/: The Eastern New England pronunciation of "tar" was observed by the experimenter in talking to subjects from Oxford, Mass. As described by Kurath and McDavid, these dialects have separate phonemes for the vowels of "add," "tar," and "odd."

f. /a/-/ɔ/: The Ontario dialect does not distinguish between the vowels of "cot" and "caught." (Avis, 1966) The speakers of Oxford, Massachusetts, also do not make this distinction. This latter fact is based on the unanimous opinion of Massachusetts subjects, that the words "cot" and "caught" contain the same vowel. This is consistent with Kurath and McDavid, although it is not possible to determine from their study what the pronunciation of this contrast is for the particular area of Oxford, Massachusetts.

2. A Summary of Inter-group Differences.

a. There are no significant differences between

---

phoneticians were asked to judge the phonetic quality of the vowels. Based on these judgements, the degree of raising of /ae/ and fronting of /a/ was determined for each subject. Our study indicated that the Ontario and Rochester groups could easily be distinguished by degree of fronting of /a/, but there was a considerable over-lap between the two groups in raising of /ae/.

[14] Kurath and McDavid, 1961, Map 15, also p. 104, "The Vowel of Crop, Oxen, College." Also compare p. 15, "Upstate New York and Western New England," and p. 19, "The Upper South."

groups from the same immediate area (i.e. town or city). Buffalo and Rochester groups, which are from the same dialect region (Western New York), show significant differences in only three of sixteen possible cases. (cf. Tables X, XI, XII, and XVII)

b. Ontario groups favor /e/ for stimulus 29, in contrast with other groups.

c. One Virginia group, one trial of Massachusetts, and both trials of Rochester A, show an area of split /E/e/.

d. Buffalo and Rochester groups contrast with other groups in that they show poor performance for the /E/-/ae/ categorization of Test IIA.

e. Buffalo and Rochester /E/-/ae/ boundaries for Test IIA contrast with those of other groups, in that they tend to divide the stimuli into fronted (/ae/) and backed (/E/) sounds, rather than into raised and lowered, as with other groups.

f. Buffalo and Ontario contrast in position of the /ae/-/a/ boundary. Buffalo shows a boundary more fronted than Ontario.

g. Massachusetts shows a center for /ar/, not found in Rochester A data.

h. All groups had difficulty categorizing /a/-/ɔ/. Virginia was the best able to make this categorization.

Massachusetts shows contradictory data, in that these subjects showed /a/ and /ɔ/ areas for both Test IB and Test IIB. In both cases, however, there was reason to suspect that these areas might be artifacts of the testing situation. Buffalo did not show well defined areas for this categorization in either test.

i. Buffalo contrasts with Virginia in position of the /a/-/ɔ/ boundary. Buffalo showed a boundary more fronted than did Virginia.

j. Ontario contrasts with other groups, in having the /e/-/i/boundary higher. This is equivalent to saying that Ontario hears /e/ for a stimulus (number 29) which other groups hear as /i/. Similarly, Ontario has a higher /o/-/ʊ/ boundary and hears stimulus 28 as /o/, where other groups, (particularly Rochester A and Virginia) here /u/.

3. Correlation of Inter-group Differences and Dialect Features.

The main hypothesis of this study, that there are dialectally determined variations in vowel perception, seems strongly supported by the fact that the significant differences between groups do not occur between groups of the same dialect area, nor between two trials of the same group. Interesting and significant differences in vowel perception do occur between groups from different dialects.

While we shall attempt below to correlate specific inter-group differences with specific dialect features, it should be noted that we have in no way experimentally proved the causal relation between dialect feature and perceptual data. Our procedure is simply to note a correlation, and to venture a probable explanation.

Perhaps the most easily explained inter-group difference is the existence of a phoneme area for the vowel of "tar," for Massachusetts subjects. This is entirely consistent with what is known of this dialect. (cf. above, F.1.a, p. 87)

A second easily explained inter-group difference is the more fronted /ae/-/a/ border for Buffalo and Rochester. This too is entirely consistent with the fact that these dialects have a more fronted /a/ vowel.

The poor performance on Test IIA by Rochester A and Buffalo groups seems to be characteristic of this dialect. There are a number of ways to relate this fact to dialect. We might suppose that speakers of this region pronounce "bet" and "bat" alike. A second explanation, though not supported in the literature, seems consistent with our own observations of the quality of the vowel of "bat" in this region. For these speakers, /ae/ seems to be distinguished from /E/ in that it is more fronted, and more of a diphthong, not necessarily lower. That the /E/-/ae/ distinction is one of fronting is indicated by the observed position of this border in the data for Rochester A and Buffalo (cf. above, F.2.e, p. 88).

Preference by Ontario groups for /e/,/o/ where other groups preferred /i/,/u/, respectively, was an unexpected but apparently reliable tendency. The most probable explanation for this result seems to be that the Ontario /e/ and /o/ are less diphthongized than those of other groups. Our reasoning in this case is as follows. We assume here that for all dialects, /i/ is less diphthongized than /e/. Note further, that our stimuli are phonetically not diphthongized. In the ambiguous are of the F1/F2 chart, for example between /i/ and /e/, (stimulus 29), Ontario groups make the decision between /i/ and /e/ entirely on the basis of the quality of the vowel. Other groups for example Virginia and Rochester A, prefer /i/, and reject /e/ because they do not hear diphthongization. (For stimuli lower than number 29, these same groups may choose /e/ because, in spite of the lack of diphthongization of the stimuli, the quality of the vowel is sufficiently different from /i/ as to exclude this response.)

An identical argument to the one in the preceeding paragraph explains Ontario preference for /o/ in stimulus number 28. /u/ is less diphthongized than /o/ for all dialects. Ontario groups prefer /o/ on the basis of vowel quality, while other groups apparently reject /o/ because the stimulus is not diphthongized.

The split /E/e/ for Rochester A, one trial of Massachusetts, and one Virginia group, may reflect a dialectal diphthongization of the vowel of "egg." This is consistent with what is known of Virginia dialect, but it is less clear that these data reflect the dialects of Rochester A and Massachusetts subjects. It is possible that this feature does not correlate with regional dialect because it is socially determined.

The results in categorization of "cot" and "caught" are difficult to explain by dialect. Firstly, we may ask why performance on this test, for all groups, was not so good as for other tests. It is possible that, as in Test IIA for Buffalo and Rochester A, quality of the vowel is not the only distinguishing feature. This explanation is not consistent with most characterizations of this contrast in the literature. The second problem is to explain how Massachusetts subjects can categorize a distinction that they do not make in their own speech. In answer to this, it should be pointed out that their performance differs in several respects from that of Virginia. They do not show well established phoneme areas for /ɔ/ in Test IB, and the /a/-/ɔ/ boundary in Test IIB is extremely unstable. A possible explanation of their ability to make the /a/-/ɔ/ distinction is that these subjects are aware that this is a distinction made in other dialects. (The subjects from Massachusetts were made aware of the /a/-/ɔ/ distinction by the experimenter before Test IIB was administered. This was done in order to justify giving the test by explaining that some speakers make this distinction.) Knowing this, the subjects were able to place /ɔ/ in proper relation to other phoneme areas simply by acoustic discrimination, without reference to linguistic training. The fact that this is not a well-learned distinction is reflected in the pecularlaties of their performance as compared with Virginia. A third problem lies with the Buffalo groups who also have difficulty with the /a/-/ɔ/ categorization, even though it is supposed to be in their dialect.[15] This could reflect the partial loss of this distinction under influence from Ontario, or for other reasons. A second explanation is that quality of the vowel is not the feature separating /a/-/ɔ/.

In spite of the difficultires of explaining the data from Test IIB, it seems clear that whatever the reasons, Virginia groups show better defined areas and borders for this contrast. Secondly, the Buffalo group showed a significant difference in position of the /a/-/ɔ/ boundary, in relation to Virginia. This experimentally determined fact that the Virginia boundary was more backed than the Buffalo boundary is consistent with what is known of these dialects.

---

[15] cf. Kurath & McDavid, 1961, pp 6,7.

IV. CONCLUSIONS

A. PERFORMANCE

The first conclusion we may draw from this study is that it is feasible to test large numbers of untrained subjects of average ability, in less than ideal testing situations. In spite of problems of inattention, poor acoustics, and pressures of time which did not allow repeated testing, it is possible to get clear phoneme centers and phoneme borders. Performance is improved by using female subjects, and by using forced choice testing format, rather than an open-ended format.[16]

The stimuli used in these tests varied in formant peak frequency in intervals of 40-60 Hz. In view of the several well defined phoneme boundaries, and several instances of good individual performance, observed in the data, it seems that this interval of 40-60Hz was not too small. On the contrary, a smaller interval might have been desirable. This would have necessitated more stimuli per test, and therefore, longer tests. However, it might have given more information as to the exact position of phoneme boundaries.

---

[16] It might be contended that females, since they are generally considered to have different vowel quality from males, (cf. Joose, 1948, p. 64) would have different positions for phoneme boundaries. We tabulated male and female separately, for each group for Tests IIA and IIC. There was no reliable or significant difference in boundary for different sexes of the same group.

Evidence that the forced choice version of Test IB gave better performance than the open-ended version is given in Table V. The Rochester A and Massachusetts groups show the best performance on Test IB.

As discussed above (p. 10), there is some reason to believe that vowels are not perceived categorially, in the same manner as consonants. This might lead one to think that subjects would not show clear phoneme boundaries in vowel discrimination (although this is not a claim made by proponents of the non-categorial nature of vowel perception). While it must be recognized that the data of these tests does not relate directly to the question of the nature of vowel perception, or to the motor theory of speech perception, it seems worth noting that in spite of the possible non-categorial nature of vowel perception, some subjects, on some tests, showed a remarkable ability to categorize randomized stimuli, and in many cases their performance was stable over time. In this sense at least we can say that we can talk about a phoneme category for a given subject.

B. VARIABILITY

Since it was not possible to test every group twice, our measures of variability are only approximate. It is likely that in our effort to be sure that our results were reliable, we have dismissed tendencies that will prove reliable when more exact measures of group variability over time can be made.

It is apparent from the data that within a given group, there is considerable variation among subjects. Some of this variation may be due to variation in dialect, and the experimental design could be improved by closer controls on the spoken dialect of subjects, to assure that a group of subjects speaks as nearly as possible one dialect. There is reason to believe that even when subjects speak with one dialect, they may perceive differently.[15] For whatever the reason, few group percentages of response are over 90%, and the tests while indicative of group tendencies, could hardly be used to classify a specific individual as to spoken dialect (except with considerable error).

---

[15] We have already seen in our discussion of articulation and perception (p. 14), that there is reason to believe that perception of synthetic vowels reflects listening habits more than speaking habits. In a short experiment run during the time of this thesis (see p. 86, foot note 13, above) we found that while there was a tendency for a person's articulation of /a/ and /ae/ to be reflected in his categorization of these sounds, there were significant exceptions. There was at least one subject out of the ten who showed extremely raised articulation of /ae/, and perceived the /E/-/ae/ boundary as extremely low.

Further evidence of the importance of listening habits, as opposed to speaking habits, in the results of perception tests, is the fact that a common question asked by subjects after taking the test was: "Who was that speaking?" This question reflects first of all the fact that the subject was not aware that the stimuli were synthetic, and secondly, that in perceiving sounds, he was concerned with who he was listening to. Thus one important variable in categorization behavior, besides dialect, may be attitude of the subject toward the person he imagines to be speaking the stimuli.

As concerns variation of group tendency over time, our data show that this variation is not so great as to obscure inter-group differences.

Finally, we considerably improved the accuracy of our measurement by using tests of the Format of IIIC, with a single row of stimuli heard four times. It is important to realize, however, that such a test is only useful after more general investigation of inter-group differences is made by tests of the type of Series I and Series II. For example, important information about the nature of the /E/-/ae/ boundary would not have been available, if we had used only a single line of stimuli for testing this contrast.

C. PRACTICAL APPLICATIONS OF SYNTHETIC VOWEL PERCEPTION TESTS TO DIALECT STUDY

Because of the considerable inter-subject variation described above, these tests can only be used to describe group tendencies. Moreover, only certain dialect features show up clearly in the data. The existence of an /ar/ phoneme center, performance in categorizing /E/-/ae/ in Test IIA, and the positions of the /ae/-/a/ and /i/-/e/ boundaries are all features that could be used to identify the dialect tendency of a group.

We can imagine a possible experiment in which one wished to study the effects of movement into a new dialect area, upon a group of subjects. Suppose that one tested two groups of twenty subjects from Rochester, N. Y., and that both groups then moved to Ft. Erie, Canada. One hypothesizes that, due to non-linguistic (e.g. social) reasons, one group will adapt more rapidly to the surrounding dialect. One then retests the two groups one year later, and uses the data from the test as a measure of dialect tendency.

It is interesting that our data did not correlate in every case with previously established views concerning the dialects studied. It seems likely that synthetic vowel perception tests can provide a scientific method of arbitrating disagreements about the discription of a dialect. One case in point is the problem of the raised /ae/ in Buffalo and Rochester. Our results indicate that the distinctive feature of this dialect is perhaps not vowel height. A second example is the lack of diphthongization of /e/,/o/ in the Ontario dialect. If our explanation is correct, then we can attribute to synthetic vowel perception tests the capability of pin-pointing heretofore undescribed dialect features.

D. THEORETICAL VALUE OF THE STUDY OF VARIATION IN PERCEPTION OF SYNTHETIC VOWELS

All too often, for lack of a better means of description, phonological features are described in terms of "all or nothing." A dialect either has or does not have a given distinction. There are obvious advantages to a description of this sort. (The concept of langue is commonly associated with an argument for an "ideal speaker" and an "ideal phonology.") Without minimizing the importance of the concept of the ideal speaker, we should like to point out that dialectal habits, like most social habits, can and perhaps should also be described in terms of statistical tendencies. Synthetic vowel perception tests of the sort used in this study provide a natural means of describing a tendency towards, for example, fronting of /ae/, or the loss of a particular distinction. With these tests such tendencies can be converted to percentages in a very natural and rational way.

We had assumed at the outset of our investigation that there was no need to ask subjects to categorize a distinction that they claim not to make. Our data indicates that this is a false assumption. While we were able in this study to differentiate between a group that has learned a given distinction as part of their dialect, and a group which is perceiving a distinction only because asked to, the exact differences in categorization behavior between these two types of groups is material for further investigation.

It is possible that if a group distinguishes between two sounds on the basis of a feature which is not a variable in a given test (e.g. /E/-/ae/ for Buffalo and Rochester), then this group will try to distinguish the sounds on the basis of the feature they have learned to look for, and will fail to perform well. On the other hand, a group that does not make any distinction between the two sounds of a given test (e.g. /a/-/ɔ/ for Massachusetts) will not be hindered by looking for a feature which is not a variable, and can more readily perform the task demanded of them, even though they rely on non-linguistic, hearing ability, and even though their performance apparently does not reflect a linguistic ability. This is a hypothetical explanation of why Buffalo and Rochester could not perform a distinction usually attributed to their dialect, while Massachusetts performed reasonably well a distinction not in their dialect. It should be possible in further investigation to test the validity of this explanation.

BIBLIOGRAPHY

(1934) Ahrend, Evelyn.: "Ontario Speech," *American Speech* IX (1934) 136-39.

(1964) Aungst, L. F. and Frick, J. V.: "Auditory Discrimination Ability and Consistency or Articulation of /r/," *J. Speech Hearing Dis.* 29.1 (1964), 76, 85.

(1966) Avis, Walter S.: "Why a Canadian Dictionary," *Speaking of Dictionaries* No. 1 (October, 1966).

(1960) Bronstein, Arthur J.: *The Pronunciation of American English*, N. Y. (1960).

(1962) Carpenter, A., and Morton, J.: "The perception of Vowel Color in formantless complex sounds," *Language and Speech* V (1962), 205-214.

(1964) Castle, W. E.: *The Effect of Narrow-band Filtering on the Perception of Certain English Vowels*, 's Gravenhage, 1964.

(1961) Chistovich, L. A.: "Classification of Rapidly Repeated Speech Sounds," *Akusticheskii Zhurnal, 6* (1960), 392-398. (Trans. in *Soviet Physics-Acoustics* 6, (1961), 393-398.)

(1966) Chistovich, L. A.; Fant, G. and de Serpa-Leitao, A.: "Mimicry and perception of synthetic vowels," *Speech Transmission on Laboratory, Quarterly Progress and Status Report*, (1966) # 2. pp. I-1-I.18 # 3, pp. I.1-I.3 (in two parts).

(1963) Cohen, A.; Slis, I. H. and 't Hart, J.: "Perceptual Tolerances of Isolated Dutch vowels," *Phonetica* 9 (1963), 65-78.

(1961) Cooper, F. S.: "Speech Synthesizers," *Proceedings of the Fourth International Congress of Phonetic Sciences*, Helsinki, (1961).

(1964) Delattre, Pierre: "Comparing the vocalic features of English, German, Spanish, and French," *IRAL* II (1964), 71-97.

(1951) Delattre, P.; Liberman, A. M. and Cooper, F. S.: "Voyelles synthetiques à deux formants et voyelles cardinales," Le Maitre Phonetique 6 (1951), 30-37.

(1952) Delattre, P.; Liberman, A. M.; Cooper, F. S. and Gerstman, L. J.: "An experimental Study of the acoustic determinants of vowel color," Word 8 (1952), 195-210.

(1965) Denes, P. B.: "On the Motor Theory of Speech Perception," PCP V (1965), 252-258.

(1960) Fant, G.: Acoustic Theory of Speech Perception, 's Gravenhage, (1960).

(1962) Fant, G.: "Descriptive Analysis of the Acoustic Aspects of Speech," Logos 5 (1962), 71-77; in Lehiste, 1967, 93-108.

(1967) Fant, G.: "Sound, Features and Perception," Speech Transmission Laboratory, Quarterly Progress and Status Report, (1967), 1-14.

(1964) Fant, G. and Risberg, A.: "Auditory Matching of vowels with two formant synthetic sounds," Speech Transmission Laboratory, Quarterly Progress and Status Report 4 (1963) II.7-II.11.

(1965) Flanagan, J. L.: Speech Analysis, Synthesis, and Perception, Heidelberg, (1965).

(1966) Galunov, V. I. and Chistovich, L. A.: "On the relation between Motor Theory and the General Problems of Speech recognition," Soviet Physics-Acoustics, 11.4 (1966), 357-365.

(1952) Griffith, Rita Scot: "The Relation between Phonetic Patterns of Individuals with superior Average, and Poor articulation, and their Preferences among Controlled Speech-sound Stimuli," Ph. D. Diss., Ohio State University, 1952, Dab XVIII (1958), 694-696.

(1952) Jakobson, R.; Fant, G. and Halle, M.: Preliminaries to Speech Analysis. "The Distinctive Features and their correlates," Acoustic Lab., MIT Tech. Rep., # 13 (Cambridge, Massachusetts, 1952).

(1948) Joos, M.: *Acoustic Phonetics*, Language Monographs, # 23, 1948.

(1966) Kozhevnikov, V. A. and Chistovich, L. A.: *Rech' Artikuliatsia i vospriiatie*. Moscow-Leningrad, 1965. (Trans. in *Speech; Articulation and Perception*.) Washington: Joint Publications Research Service, 1966, 30, 543.

(1961) Kurath, Hans and McDavid, Raven: *The Pronunciation of English in the Atlantic States*, Ann Arbor, 1961.

(1957) Ladefoged, P. and Broadbent, D. E.: "Information conveyed by vowels," *JASA* 29 (1957), 98-104; Lehiste, 326-332.

(1965) Lane, H.: "Motor Theory of Speech Perception: a critical review," *Psych. Rev*. 72 (1965) 275-309.

(1967) Lane, H.: "A Behavioral Basis for the Polarity Principal in Linguistics," *Language* 43.2 (1967) 494-511.

(1965) Laver, J. D. M. H.: "Variability in vowel perception," *L·and S.*VIII (1965), 95-121.

(1967) Lehiste, Ilse, ed.: *Readings in Acoustic Phonetics*, MIT, 1967.

(1967) Leon, P., and Nemni, M.: "Franco-canadien et francais standard: problemes de perception des oppositions vocaliques," *CJL/RCL* 12 (1957), 97-112.

(1957) Liberman, A. M.: "Some Results of Research on Speech Perception," *JASA* 29 (1957), 117-123.

(1962) Liberman, A. M., Cooper, F. S., Harris, K. S., and MacNeilage, P. F.: "A Motor Theory of Speech Perception," *Proceedings of the Speech Communication Seminar*, Stockholm: Royal Institute of Tech., 1963, D3.

(1967) Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M.: "Perception of the Speech Code," *Psych Rev*. 74.6 (1967), 431-461.

(1966) Lindner, G.: "Veranderung der Beurteilung synthetischer Vokale unter dem Einflus des Sukzessivkontrastes," *Ztschr. f. Phonetik* 19 (1966), 287-307.

(1967) ______ "Dekodierungsprozess auditiv aufgefasster Einzellaute," *Ztschr. f. Phonetik* 20 (1967), 61-80.

(1964) Lisker, L., and Abramson, A. S., "A Cross-Language Study of Voicing in Initial Stops: Accoustical Measurements," Word 20 (1964), 384-422.

(1967) Lorge, B.: "A Study of the Relationship Between Production and Perception of Initial and Intervocalic /t/ and /d/ in Individual English Speaking Adults," *Status Report on Speech Research*, Haskins Laboratories, SR-9 (1967), 3.1-3.18.

(1967) Menyuk, P.: "Children's perception of a set of vowels," *Quart. Progress Rep., Research Lab. of Electronics*, M.I.T. (1967), No. 84, 252-257.

(1962) Morton, J., and Carpenter, A.: "Judgement of vowel color of natural and artificial sounds," *L. and S.* V (1962), 190-204.

(1963) ______ "Experiments relating to the perception of formants," *JASA* 35 (1963), 475-480.

(1965) Obrecht, D. H.: "Toward Automated Phonological Mapping," *Linguistics* 12 (1965), 21-35.

(1963) Oeken, F. W.: "Kritisches zur Formanttheorie der Vokale," *Phonetica* 10 (1963), 22-33.

(1961) Peterson, G. E.: "Parameters of Vowel Quality," *J. Speech Hearing Res.* 4 (1961), 10-29.

(1952) Peterson, G. E., and Barney, H. L.: "Control Methods Used in a Study of the Vowels," *JASA* 24 (1952), 175-184; in Lehiste, 1967, pp. 118-127.

(1958) Sapon, S., and Carroll, J. B.: Discriminative Perception of Speech Sounds as a Function of Native Language," *Gen. Ling.* 3 (1958), 62-73.

(--) Scholes, R. J.: "Categorial Responses to Synthetic Vocalic Stimuli by Native Speakers of Various Foreign Languages," Final Report, Indiana Language Program Grant No. 40-293-42 (unpublished paper).

(1967) ______ "Phoneme Categorization of Synthetic Vocalic Stimuli by Speakers of Japanese, Spanish, Persian, and American," *L. and S.* 10 (1967) 46-68.

(1968) ______ "Perceptual Categorization of Synthetic Vowels as a Tool in Dialectology and Typology," L. and S. 11 (1968), 194-207.

(1967) Shankweiler, D., and Studdert-Kennedy, M.: "Identification of Consonants and Vowels Presented to Left and Right Ears," *Quart. J. of Exper. Psych.* 19(1967), 59-63.

(1956) Siegel, Sidney: *Nonparametric Statistics*, McGraw-Hill, 1956.

(1966) Singh, S.: "Cross-language Study of Perceptual Confusion of Plosive Phonemes in Two Conditions of Distortion," *JASA* 40.3(1966), 635-656.

(1966) Singh, S., and Black, J.: "Study of 26 Intervocalic Consonants as Spoken and Recognized by Four Language Groups," *JASA* 39.2(1966), 372-387.

(1963) Stevens, K.N., Ohman, S., and Liberman, A.: "Identification and Discrimination of Rounded and Unrounded Vowels," *JASA* 35(1963), 1900A.

(1963) Studdert-Kennedy, M., Liberman, A., and Stevens, K.N.: "Reaction Time to Synthetic Stops and Vowels at Phoneme Centers and Phoneme Boundaries," *JASA* 35(1963), 1900A.

(1947) Thomas, Charles K.: *An Introduction to the Phonetics of American English*, New York, 1947.

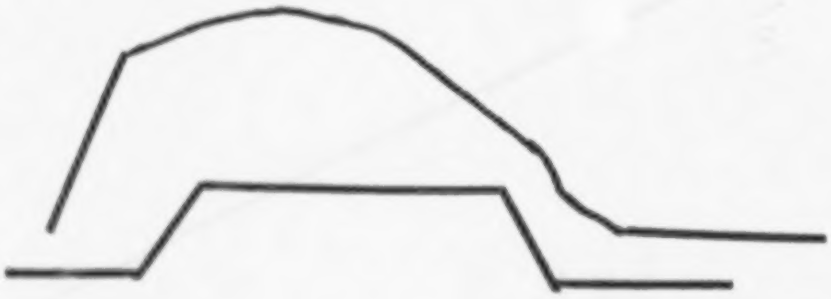
(1957) Wise, C.M. : *Applied Phonetics* , New York, 1957.

APPENDIX A-1

Settings of EVA used to produce the synthetic vowel stimuli

a. Pitch (F0)[1] :

b. Amplitude (A0)[1] :



c. Other parameters[2] :

F1 - changed for each stimulus item according to Appendix A-2

F2 - changed for each stimulus item according to Appendix A-2

F3 - dial set at 3500Hz

F4 - dial set at 4500Hz

F5 - dial set at 5000Hz

Nasal Board - dial set at zero amplitude

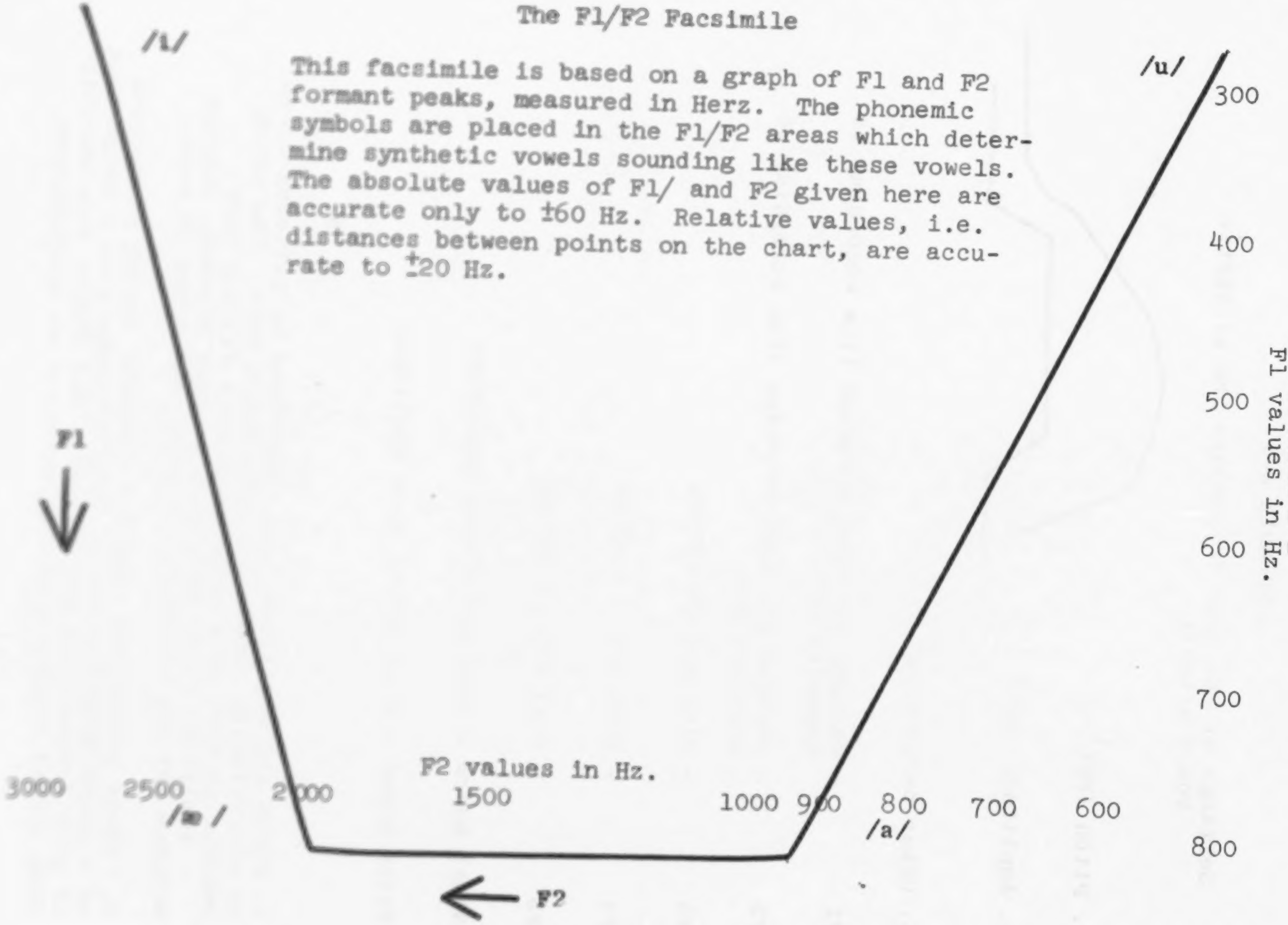
Noise Board - Dial set at zero amplitude

---

[1] Pitch and amplitude were controlled by patterns drawn in electrically conductive ink, shown here. The pitch contour is that of a male voice with falling tone.
Amplitude rises from zero to near maximum, remains maximum for approximately .3 sec., and drops to zero.

[2] These parameters remained constant during production of a given stimulus item. Dial settings give a rough idea of peak frequencies for formants, but differ considerably from actual formant peaks as measured on spectrograms.

APPENDIX B-1

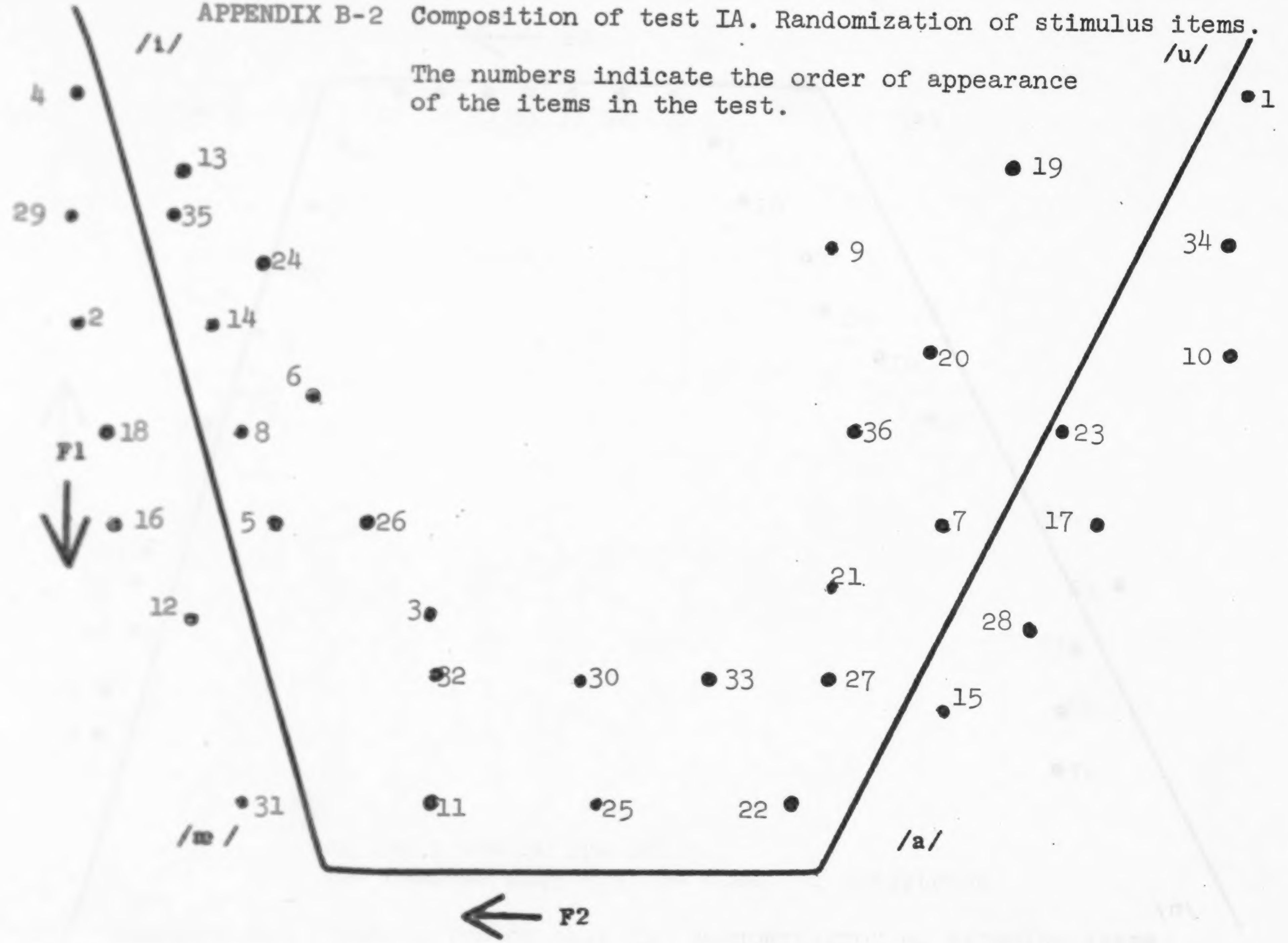


The F1/F2 Facsimile

This facsimile is based on a graph of F1 and F2 formant peaks, measured in Herz. The phonemic symbols are placed in the F1/F2 areas which determine synthetic vowels sounding like these vowels. The absolute values of F1/ and F2 given here are accurate only to ±60 Hz. Relative values, i.e. distances between points on the chart, are accurate to ±20 Hz.

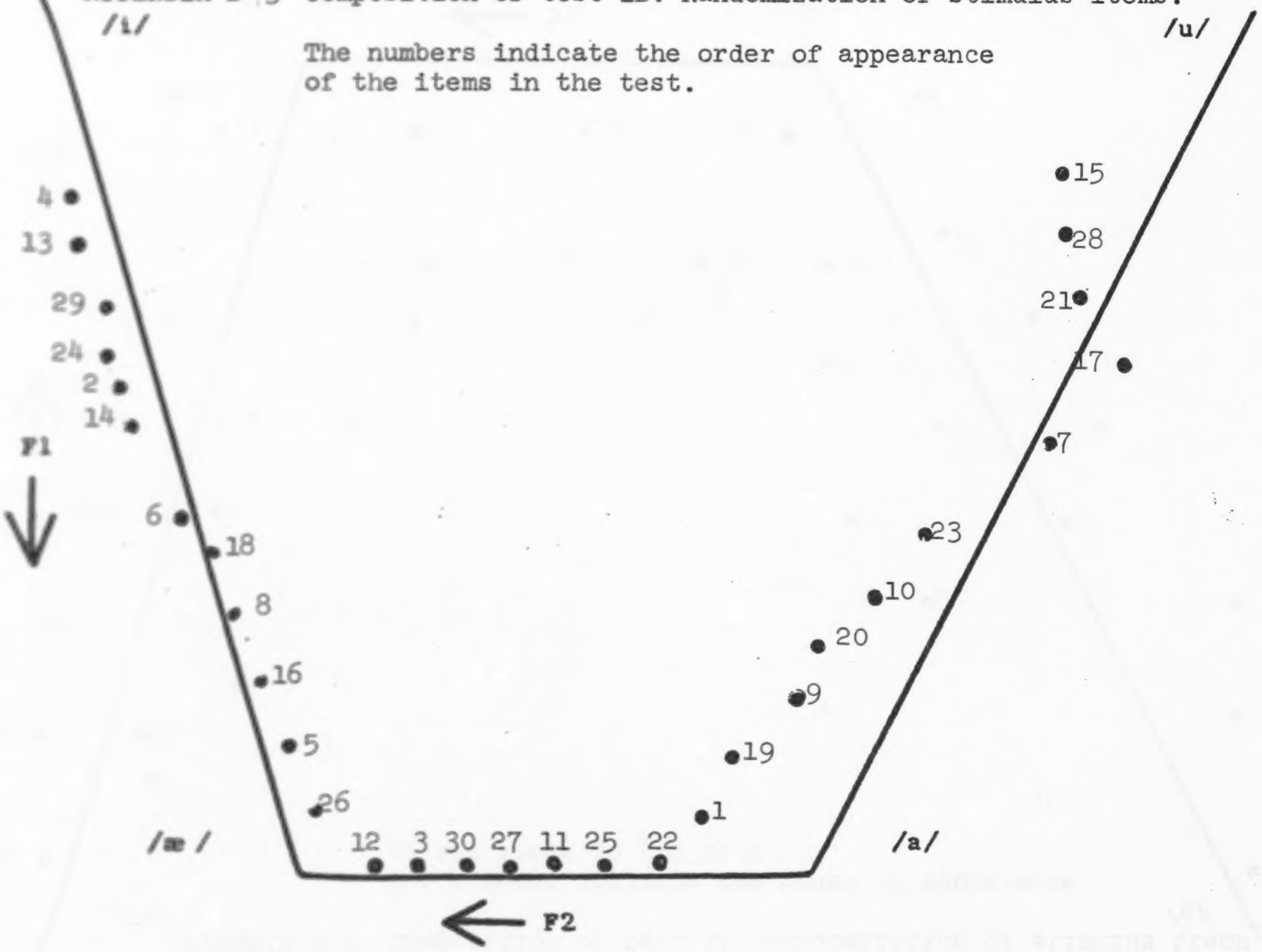
APPENDIX B-2 Composition of test IA. Randomization of stimulus items.

The numbers indicate the order of appearance of the items in the test.



APPENDIX B-3

APPENDIX B-3 Composition of test IB. Randomization of stimulus items.

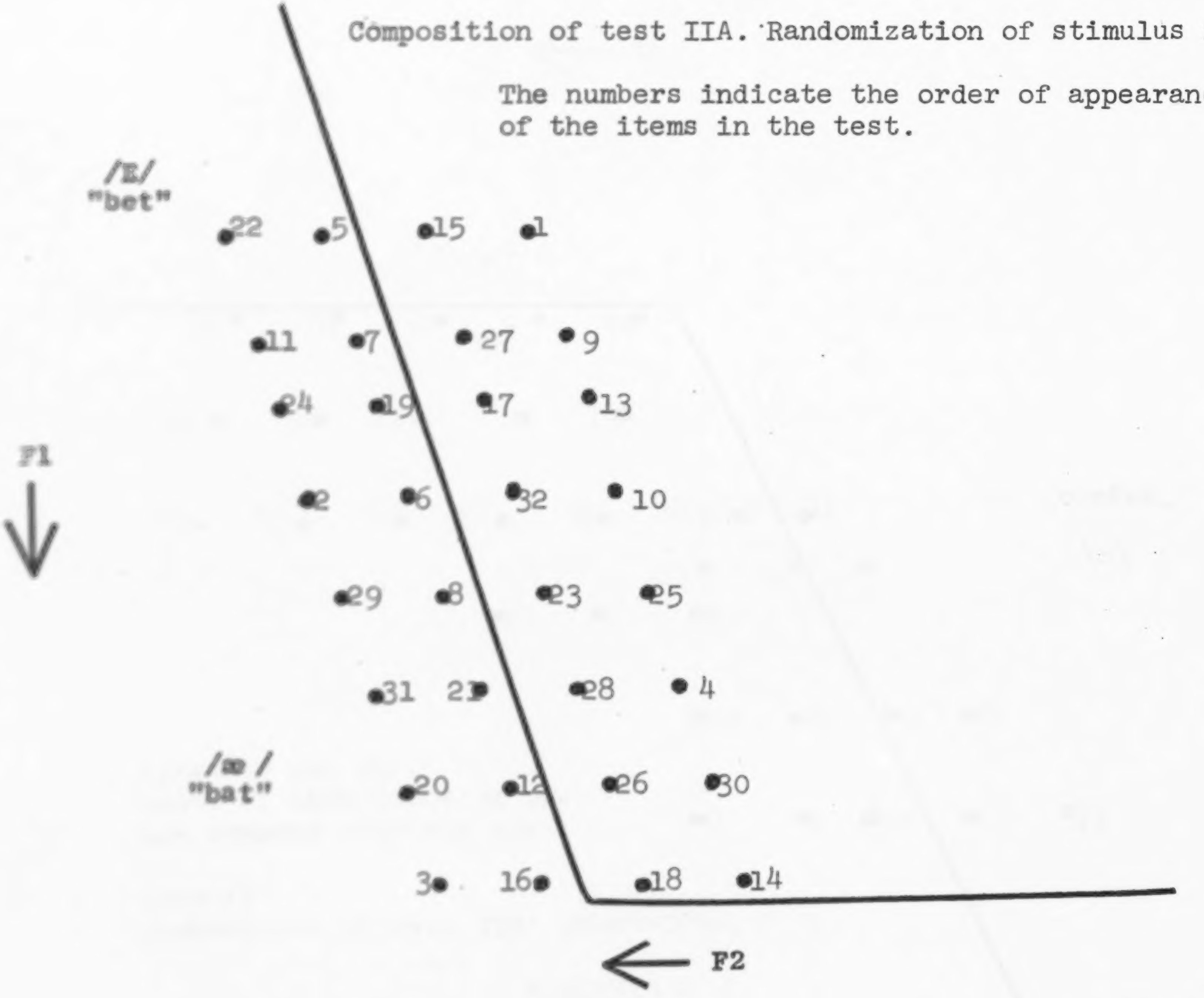


The numbers indicate the order of appearance of the items in the test.

APPENDIX B-4

Composition of test IIA. Randomization of stimulus items

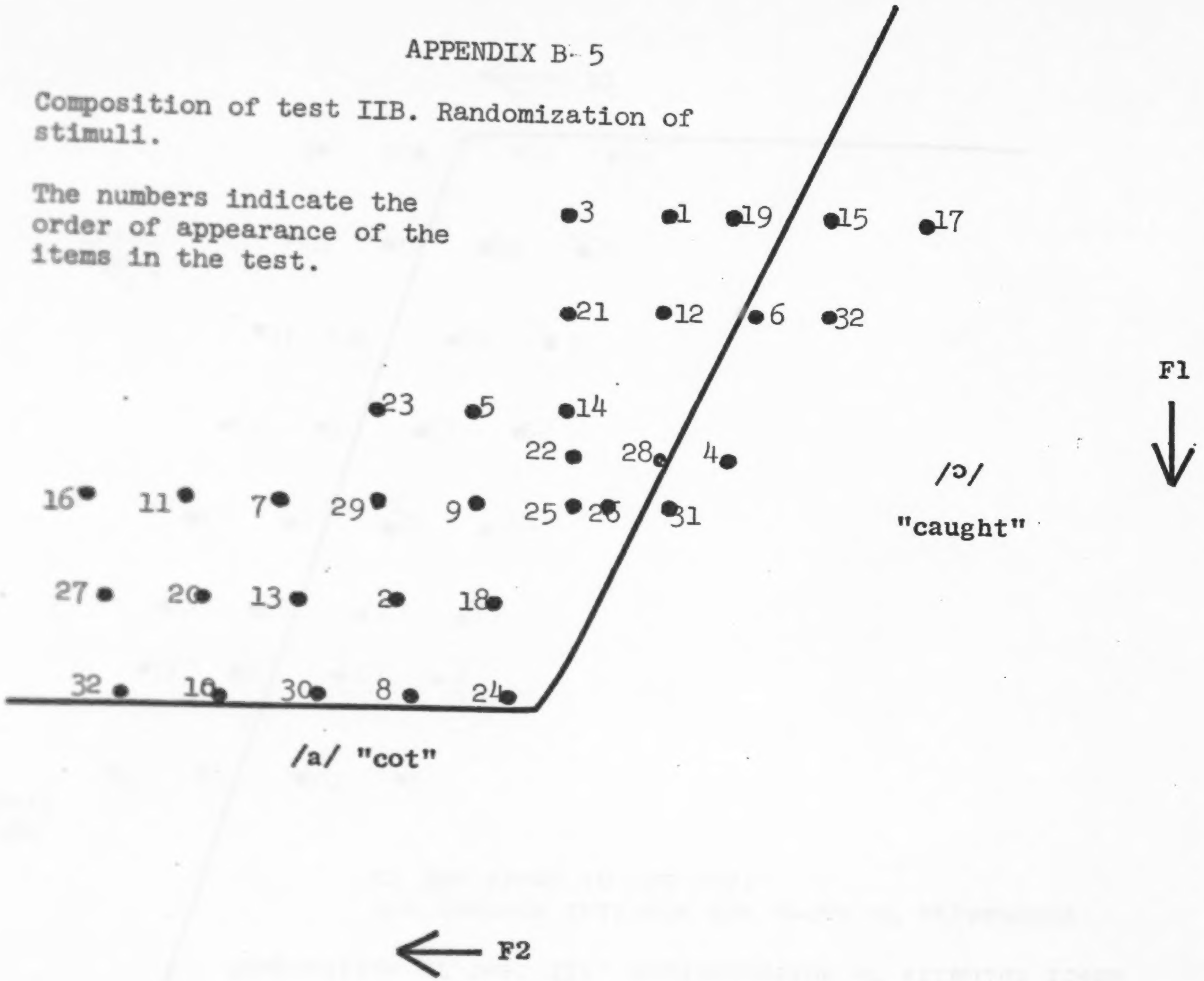
The numbers indicate the order of appearance of the items in the test.


/E/
"bet"
22
5
15
1
11
7
27
9
24
19
17
13
F1
2
6
32
10
29
8
23
25
31
21
28
4
/æ/
"bat"
20
12
26
30
3
16
18
14
F2

APPENDIX B-5

Composition of test IIB. Randomization of stimuli.

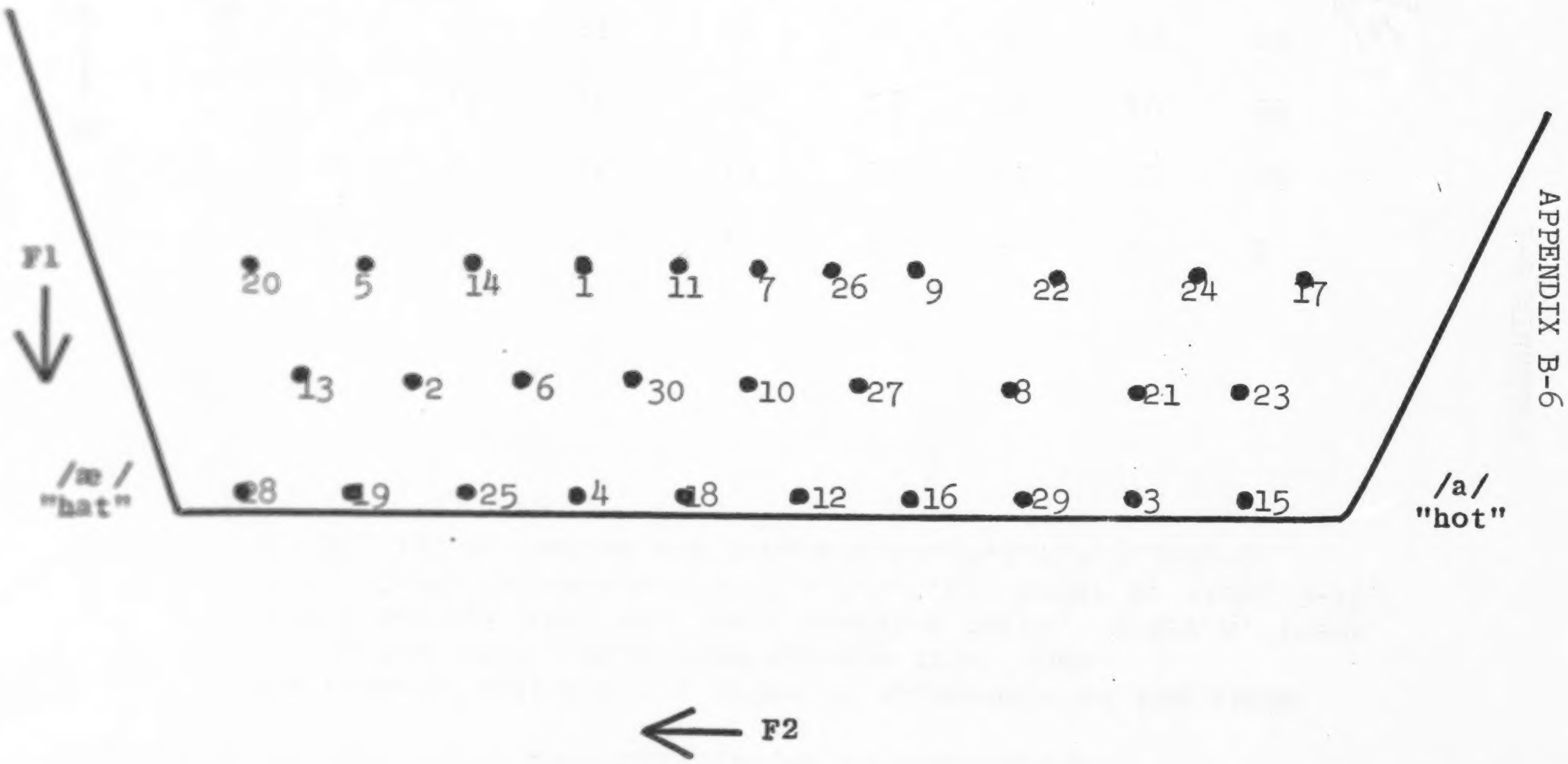
The numbers indicate the order of appearance of the items in the test.



APPENDIX B- 6

Composition of test IIC. Randomization of stimulus items.

The numbers indicate the order of appearance of the items in the test.

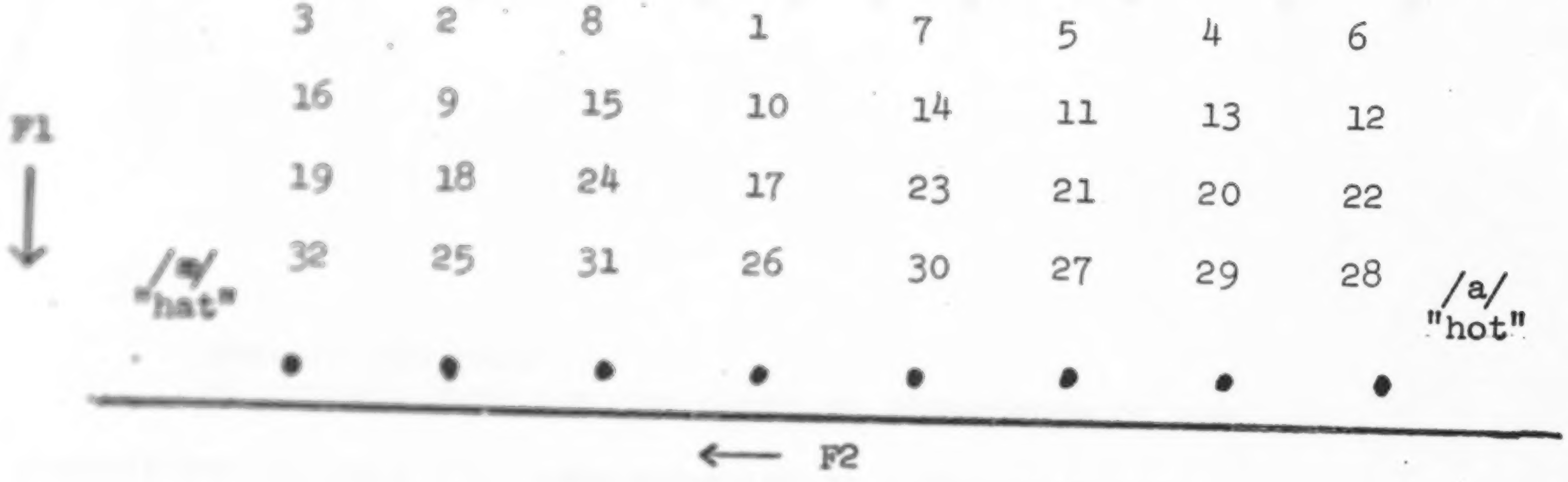

F1
/æ /
"hat"
/a/
"hot"
F2
20
5
14
1
11
7
26
9
22
24
17
13
2
6
30
10
27
8
21
23
28
19
25
4
18
12
16
29
3
15

APPENDIX B- 7

Composition of test IIIC, Randomization of stimulus items.

The numbers indicate the order of appearance of the items in the test. Each item appears four times.

Two orderings are used, each repeated twice. Order A, items 1-8, 17-24, is random.(3,2,8,1,7,5,4,6) Order B, items 9-16, 25-32, is non-random and progressive.(8,1,7,2,6,3,5,4)



3	2	8	1	7	5	4	6
16	9	15	10	14	11	13	12
19	18	24	17	23	21	20	22
32	25	31	26	30	27	29	28

APPENDIX C-1

Questionnaire given to all subjects to determine to what degree they were native to the region in which they were living when tested.

Name ....

Date of Birth ....

Place of Birth: City.........................State....

Father's Place of Birth:City..................State....

Mother's Place of Birth:City..................State....

What Grade are you in?................................

Where did you attend elementary school? City...........

.............................State....................

Where did you attend high school? City................

.............................State....................

APPENDIX C-2

Written instructions used in administering tests.

NOTE: Shown here are instructions for Tests IB, IIA. In the instructions given here for Test IB, subjects are asked to write "a one-syllable English word." In a second version of this test, subjects were asked to choose one of nine words given at the top of the answer sheet. (cf. Appendix C-3)

The same instructions were used for Tests IA, IB.

Except for changes in word categories (here "bet," "bat") number of items, the same instructions were used for all tests in Series II and III.

In most cases the instructions were read by a teacher or other adult familiar to the subjects and not by the experimenter.

* * *

INSTRUCTIONS FOR TEST IB:

You are going to hear a series of vowel-like sounds. When you hear the sound, please write on your answer sheet a one-syllable English word containing the same vowel sound.

You may use the same word for more than one item, if it sounds the same.

There are ..30.. items in the test. You will have to use some imagination to hear some of these as vowels. If you cannot think of any riming word, you may put an X, but try to put a word, even if it is a guess. If you miss an item, put an X, and go to the next.

Are there any questions? Remember that your answers should be of one syllable, and you may use the same word for more than one item.

INSTRUCTIONS FOR TEST IIA:

You are going to hear a series of vowel-like sounds. Some will sound like the vowel of .bet... others more like the vowel of .bat.. If the sound is more like .bet.. check the column on your sheet under .bet..., if bat, under .bat..

There are ..3[illegible].. items on the test. You will have to use some imagination to hear some of these sounds as vowels. If you cannot decide between .bet.. and .bat... you may guess or leave the item blank.

APPENDIX C-3

Sample answer sheet for IB.

NOTE: In one version of Test IB, subjects were not given the nine word categories shown here, but were asked to write "a one-syllable English word." In a second version of the test they were asked to write one of the nine words at the top of the answer sheet.

A similar answer sheet was used for Test IA.

NAME TEST NO. DATE

EAT ATE EGG ADD ODD TAR TORE TWO TOE

1. ....
2. ....
3. ....
4. ....
5. ....

6. ....
7. ....
8. ....
9. ....
10. ....

11. ....
12. ....
13. ....
14. ....
15. ....

16. ....
17. ....
18. ....
19. ....
20. ....

21. ....
22. ....
23. ....
24. ....
25. ....

26. ....
27. ....
28. ....
29. ....
30. ....

APPENDIX D-1

Description of subjects: I.Q.[1] and Dialect Background

Group Buffalo A

Subject number	Dialect Background[2]
1	-
2	native
3	-
4	native
5	native
6	-
7	-
8	-
9	non-native
10	-
11	native
12	-
13	-
14	native
15	non-native
16	-
17	-
18	native
19	native
20	native

1. Mean I.Q. for the group, based on the Lorge-Thorndike test, is 105.

2. From the questionnaire.
Native is defined as being born in the region and having both parents born in the region, and being raised in the region. Non-native indicates birth outside the region. Unmarked are subjects born and raised in the area but with one or more parents non-native.

APPENDIX D-2

Description of subjects: Sex, I.Q.[1], and Dialect Background

Group Buffalo B

Subject number	Sex	Dialect Background[2]
1	F	non-native
2	F	native
3	F	native
4	F	-
5	F	-
6	F	-
7	F	-
8	F	native
9	F	native
10	F	native
11	M	-
12	F	native
13	M	native
14	F	non-native
15	F	native
16	F	-

1. Mean I.Q. for the group, based on the Lorge-Thorndike test, is 116.

2. From the questionnaire.
Native is defined as being born in the region and having both parents born in the region, and being raised in the region. Non-native indicates birth outside the region. Unmarked subjects are born and raised in the area but with one or more parents non-native.

APPENDIX D-3

Description of subjects: Sex, I.Q.[1], and Dialect Background

Group Clarence

Subject number	Sex	I.Q.	Dialect Background[2]
1	F	113	-
2	F	91	native
3	M	101	-
4	M	112	native
5	F	-	non-native
6	M	109	non-native
7	F	127	native
8	M	109	native
9	F	107	native
10	M	114	-
11	M	114	native

---

[1] Mean I.Q. for the group, based on the Otis test, is 110.

[2] From the questionnaire.
Native is defined as being born in the region and having both parents born in the region, and being raised in the region. Non-native indicates birth outside the region. Un marked subjects are born and raised in the area but with one or more parents non-native.

APPENDIX D-4

Description of subjects: Sex[1] and Dialect background

Group Rochester A

Subject number	Dialect Background[2]
1	native
2	-
3	-
4	-
5	native
6	-
7	non-native
8	native
9	-
10	native
11	native
12	native
13	-
14	native
15	-
16	native
17	native
18	non-native
19	native
20	-

[1] All subjects female

[2] See Appendix D-1, ft. nt. 2

APPENDIX D-5

Description of subjects: Sex, I.Q.[1], and Dialect Background

Group Ontario A

Subject number	Sex	I.Q.	Dialect Background
1	F	111	-
2	F	103	non-native
3	F	106	-
4	F	107	-
5	M	116	-
6	M	111	native
7	F	117	-
8	F	113	non-native
9	M	121	native
10	M	107	-
11	M	103	-
12	M	122	-
13	F	114	-
14	F	123	-
15	F	109	non-native
16	F	123	non-native
17	F	121	native
18	F	110	-
19	M	112	-
20	F	106	-
21	M	107	-
22	F	119	native
23	-	106	-
24	M	104	-
25	M	109	native
26	F	115	-
27	M	119	native
28	F	104	non-native

[1] Mean I.Q. for the group, based on the Otis test, is 112.

[2] See Appendix D-1, ft. nt. 2.

APPENDIX D-6

Description of subjects: Sex, I.Q.[1], and Dialect Background

Group Ontario B

Subject number	Sex	I.Q.	Dialect Background[2]
1	M	102	-
2	F	89	-
3	M	107	-
4	M	101	native
5	M	112	native
6	M	102	native
7	M	103	native
8	M	113	native
9	M	94	-
10	M	108	-
11	M	93	-
12	M	109	non-native
13	M	-	-
14	M	112	-
15	M	105	native
16	M	108	-
17	M	108	-
18	F	113	-
19	M	88	-
20	M	105	non-native
21	M	102	non-native
22	M	103	non-native
23	F	103	native
24	M	109	-

1. Mean I.Q. for the group, based on the Otis test, is 104.

2. see Appendix D-1, ft. nt. 2.

APPENDIX D-7

Description of subjects: Sex, IQ.[1], and Dialect Background

Group Virginia A

Subject Number	Sex	I.Q.	Dialect Background[2]
1	F	109	native
2	F	119	native
3	M	-	native
4	M	120	native
5	F	96	-
6	M	123	native
7	F	-	-
8	F	-	non-native
9	F	117	native
10	F	111	native
11	F	114	native
12	M	116	-
13	F	113	native
14	M	132	-
15	F	118	native
16	F	-	-
17	M	143	non-native
18	M	122	-
19	M	127	native
20	M	109	native
21	M	126	native
22	F	131	-
23	M	109	-
24	M	134	non-native
25	F	-	-
26	F	-	native
27	F	97	native
28	F	127	non-native
29	M	126	native
30	F	124	native

1. Mean I.Q. for the group, based on the California Mental Maturity test, is 119.
2. See Appendix D-1, ft. nt. 2.

APPENDIX D-8

Description of subjects: Sex, I.Q.[1], and Dialect Background

Group Virginia B

Subject Number	Sex	I.Q.	Dialect Background[2]
1	F	100	non-native
2	F	111	native
3	F	-	non-native
4	F	123	-
5	M	130	native
6	F	117	native
7	F	134	native
8	F	-	non-native
9	M	100	non-native
10	F	115	-
11	M	115	native
12	F	133	non-native
13	F	117	native
14	F	-	non-native
15	M	-	-
16	M	111	native
17	M	125	native
18	F	-	non-native
19	M	110	native
20	M	101	-
21	M	115	non-native
22	F	105	-
23	F	117	native

1. Mean I.Q. for the group, based on the California Mental Maturity test, is 116.
2. see Appendix D-1, ft. nt. 2.

APPENDIX D-9

Description of subjects: Sex[1] and Dialect Background

Group Massachusetts

Subject Number	Dialect Background[2]
1	native
2	native
3	native
4	-
5	native
6	-
7	-
8	native
9	-
10	native
11	-
12	-
13	native
14	native
15	native
16	non-native
17	native
18	-
19	-
20	-
21	-

1. All subjects were female

2. See Appendix D-1, ft. nt. 2.

APPENDIX E-1

Sample of Computer Readable Data Encoded from Answer Sheets

Note: The numbers refer to subjects. AE = /æ/, AA = /a/

```
OXFORD, MASS. TEST  II C  SECOND TRIAL
4    AE   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA   AA   AA
AE   AE   AE   AE   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA
AA   AA   AE   AE   AE
9    AE   AE   AE   AA   AA   AA   AE   AE   AE   AA   AA   AA   AA
AE   AE   AE   AA   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA
AA   AA   AE   AE   AE
10   AE   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA   AA   AA
AE   AE   AE   AE   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA
AA   AA   AE   AE   AE
11   AE   AE   AE   AA   AE   AA   AE   AE   AE   AE   AE   AA   AA
AE   AE   AE   AE   AE   AE   AA   AE   AA   AE   AE   AE   AE   AA
AA   AA   AE   AE   AE
12   AE   AE   AE   AA   AE   AA   AE   AE   AE   AE   AA   AA   AA
AE   AE   AA   AA   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA
AA   AA   AE   AE   AE
13   AE   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA   AA   AA
AE   AE   AE   AE   AE   AE   AA   AE   AA   AE   AE   AE   AE   AE
AA   AA   AE   AE   AE
14   AE   AE   AE   AA   AA   AA   AE   AE   AE   AE   AA   AA   AA
AE   AE   AA   AA   AA   AE   AA   AE   AA   AE   AE   AA   AE   AA
AA   AA   AE   AE   AE
16   AE   AE   AE   AA   AE   AA   AE   AE   AE   AE   AE   AA   AA
AE   AE   AE   AE   AE   AE   AA   AE   AA   AE   AE   AE   AE   AE
AA   AE   AE   AE   AE
```

APPENDIX E-2

Sample of Computer Tabulation of Group Results

Note: The Computer symbols are explained in Table I of the text.

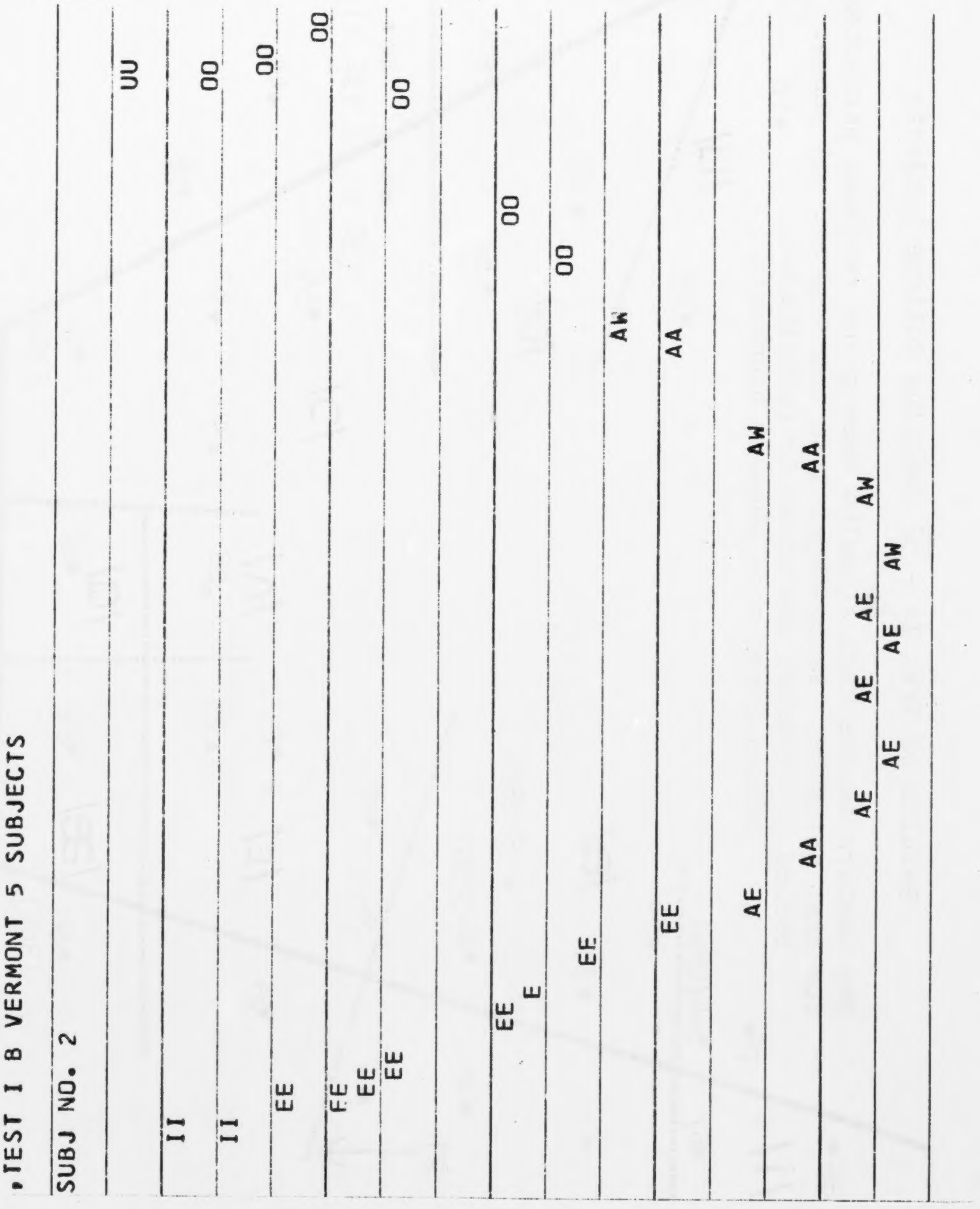
Numbers at left refer to items in order of appearance in the test. Eight subjects took this test.

OXFORD, MASS. TEST III C SECOND TRIAL

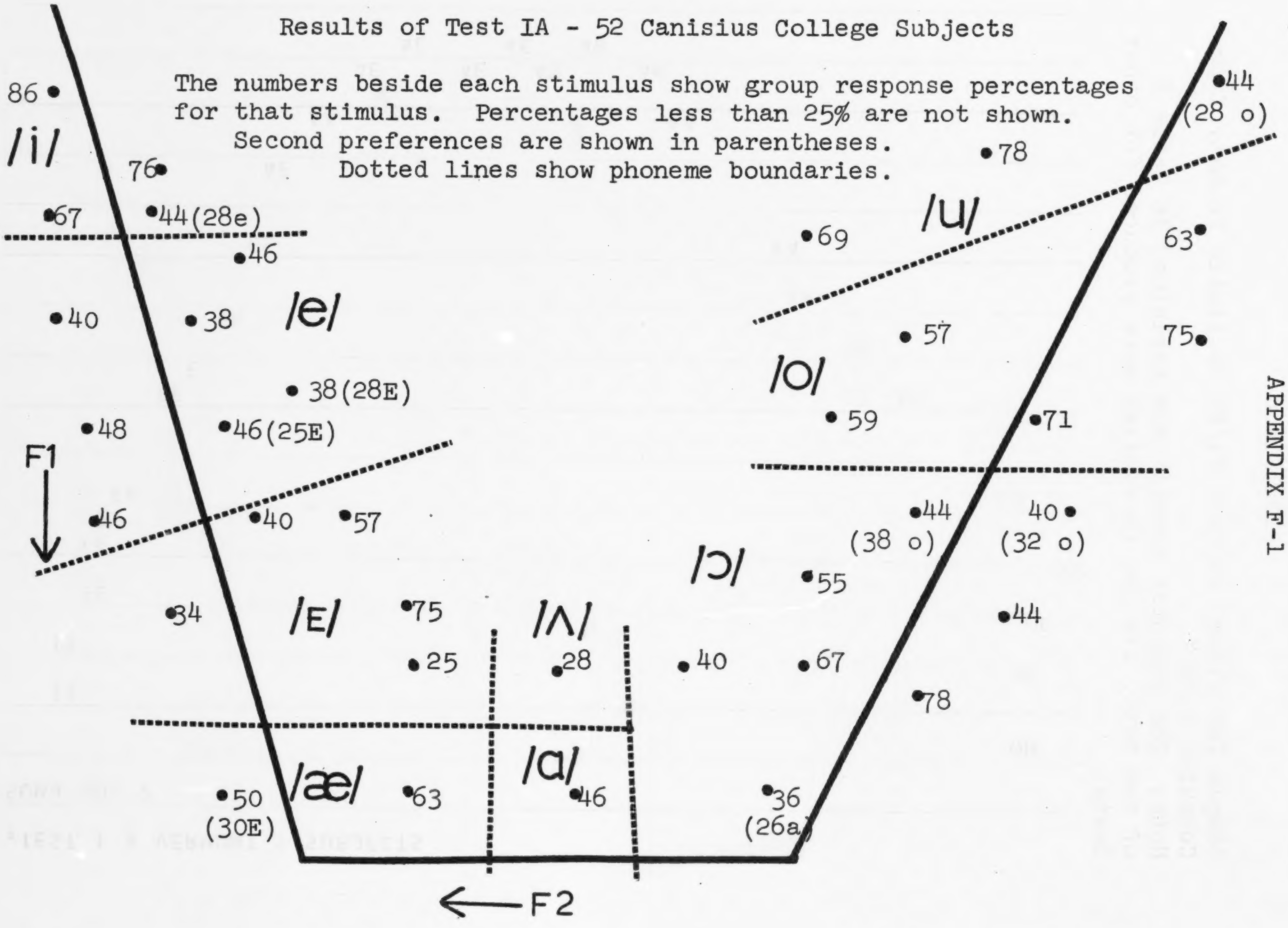
	II	I	EE	E	AE	AA	UH	AW	OO	U	UU	X
1	0	0	0	0	8	0	0	0	0	0	0	0
2	0	0	0	0	8	0	0	0	0	0	0	0
3	0	0	0	0	8	0	0	0	0	0	0	0
4	0	0	0	0	0	8	0	0	0	0	0	0
5	0	0	0	0	3	5	0	0	0	0	0	0
6	0	0	0	0	0	8	0	0	0	0	0	0
7	0	0	0	0	8	0	0	0	0	0	0	0
8	0	0	0	0	8	0	0	0	0	0	0	0
9	0	0	0	0	8	0	0	0	0	0	0	0
10	0	0	0	0	7	1	0	0	0	0	0	0
11	0	0	0	0	2	6	0	0	0	0	0	0
12	0	0	0	0	0	8	0	0	0	0	0	0
13	0	0	0	0	0	8	0	0	0	0	0	0
14	0	0	0	0	8	0	0	0	0	0	0	0
15	0	0	0	0	8	0	0	0	0	0	0	0
16	0	0	0	0	6	2	0	0	0	0	0	0
17	0	0	0	0	5	3	0	0	0	0	0	0
18	0	0	0	0	7	1	0	0	0	0	0	0
19	0	0	0	0	8	0	0	0	0	0	0	0
20	0	0	0	0	0	8	0	0	0	0	0	0
21	0	0	0	0	4	4	0	0	0	0	0	0
22	0	0	0	0	0	8	0	0	0	0	0	0
23	0	0	0	0	8	0	0	0	0	0	0	0
24	0	0	0	0	8	0	0	0	0	0	0	0
25	0	0	0	0	7	1	0	0	0	0	0	0
26	0	0	0	0	8	0	0	0	0	0	0	0
27	0	0	0	0	2	6	0	0	0	0	0	0
28	0	0	0	0	0	8	0	0	0	0	0	0
29	0	0	0	0	1	7	0	0	0	0	0	0
30	0	0	0	0	8	0	0	0	0	0	0	0
31	0	0	0	0	8	0	0	0	0	0	0	0
32	0	0	0	0	8	0	0	0	0	0	0	0

APPENDIX E-3

Sample Individual Subject F1/F2 Facsimile Format, from Computer Output.
Note: The Computer symbols are explained in Table I of the text. Similar facsimiles were produced for other tests.



APPENDIX F-1



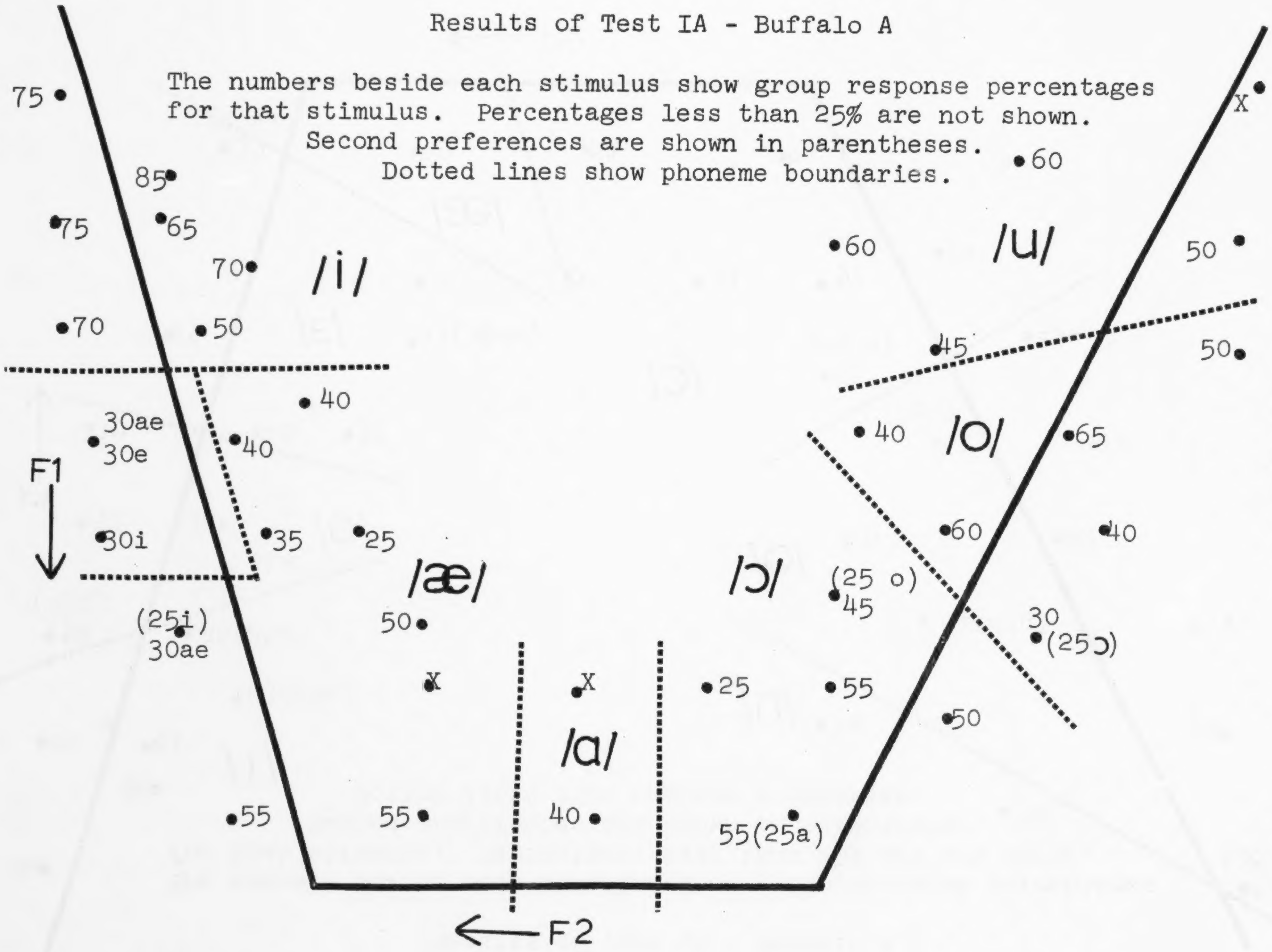
Results of Test IA - 52 Canisius College Subjects

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.
Second preferences are shown in parentheses.
Dotted lines show phoneme boundaries.

APPENDIX F-2

Results of Test IA - Buffalo A

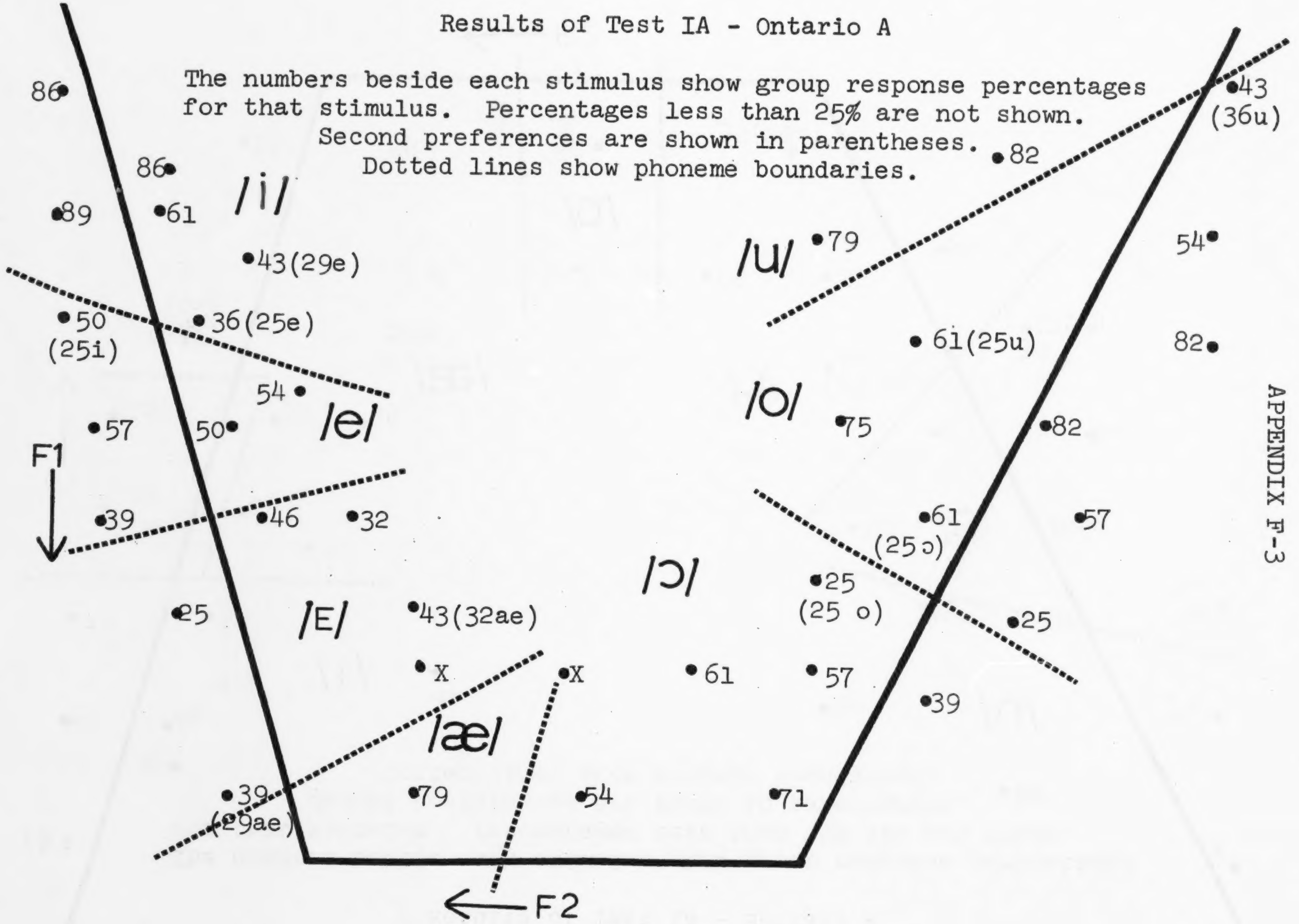
The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.
Second preferences are shown in parentheses.
Dotted lines show phoneme boundaries.



APPENDIX F-3

Results of Test IA - Ontario A

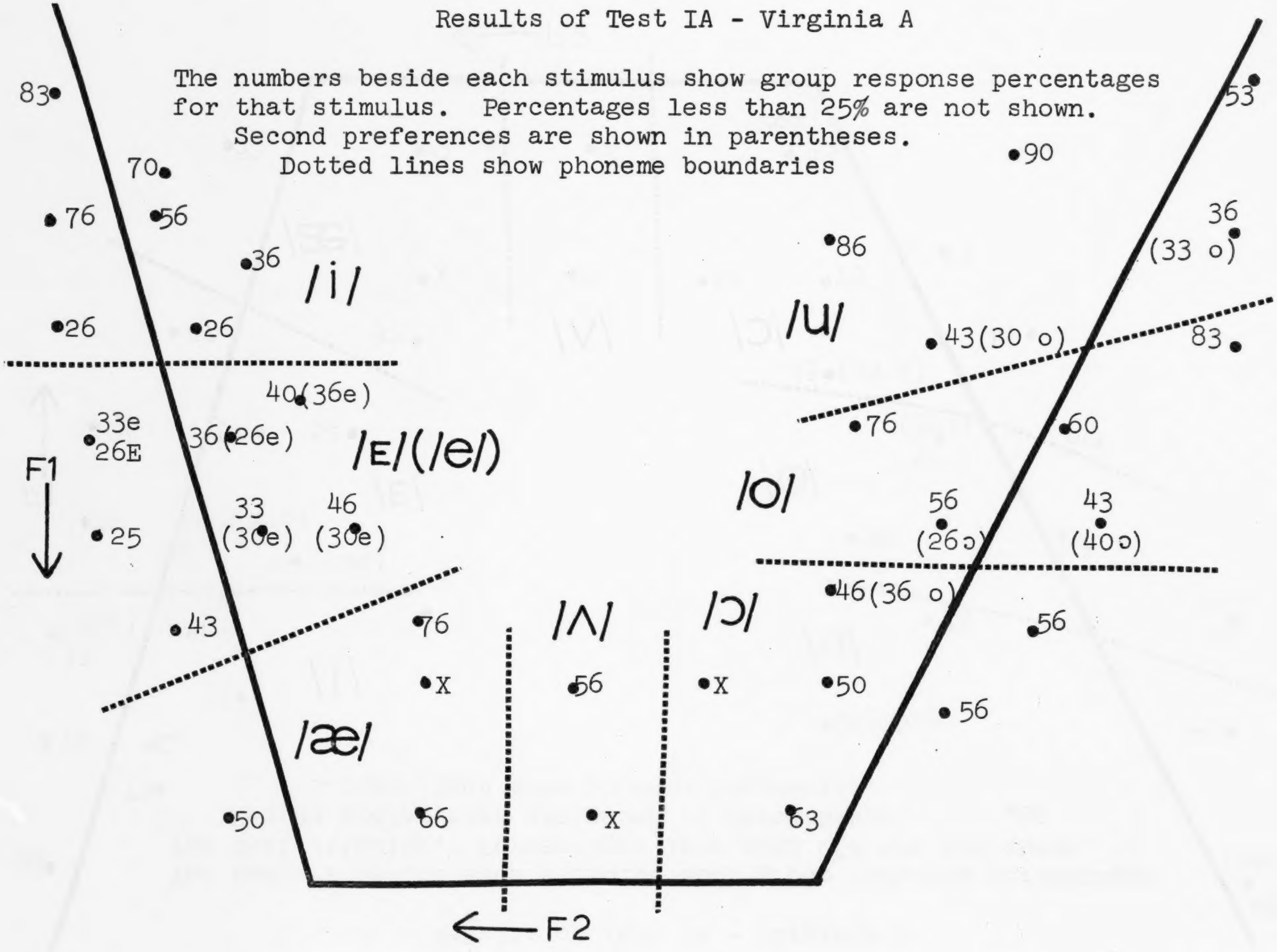
The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.
Second preferences are shown in parentheses.
Dotted lines show phoneme boundaries.



APPENDIX F-4

Results of Test IA - Virginia A

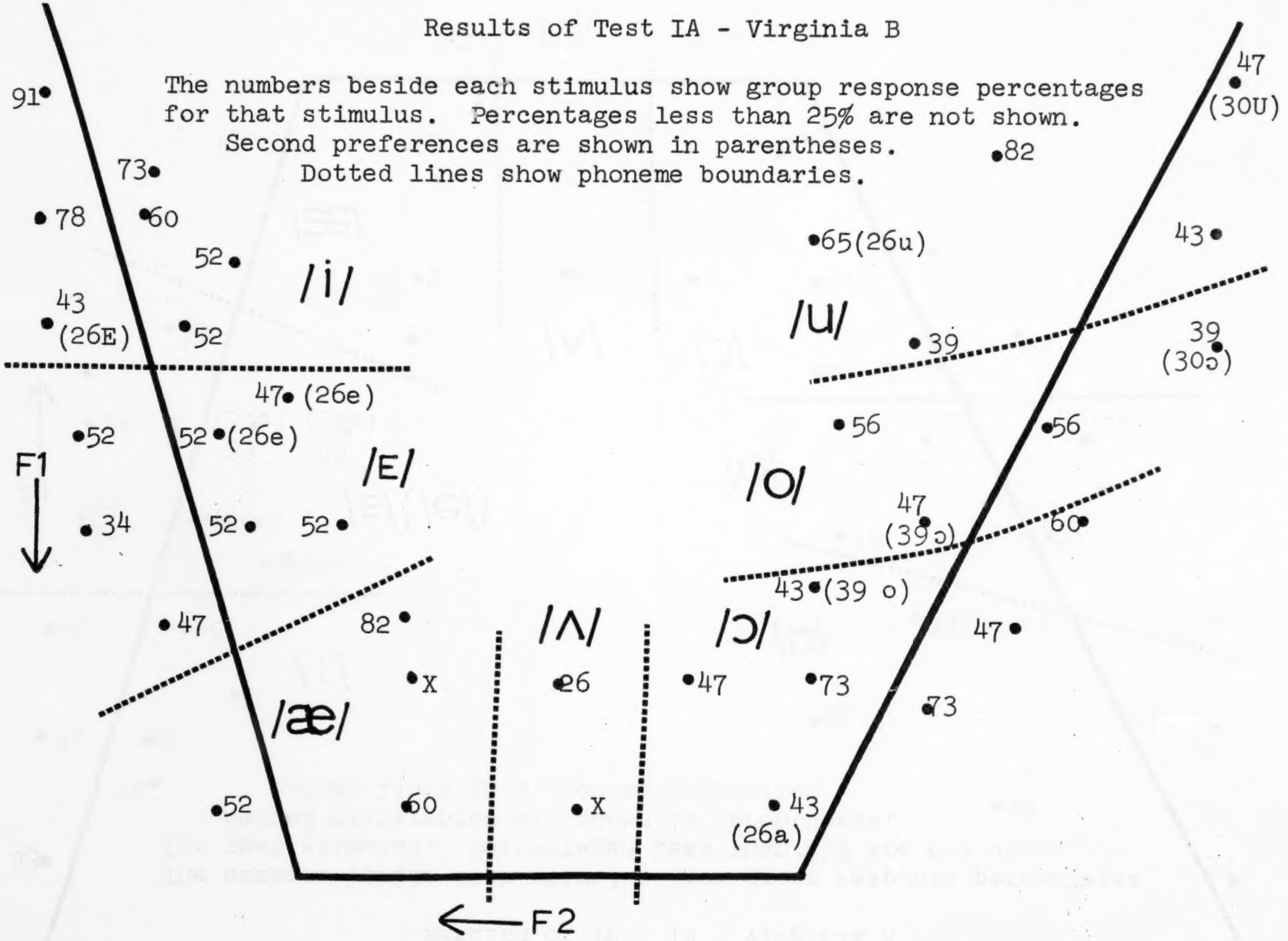
The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.
Second preferences are shown in parentheses.
Dotted lines show phoneme boundaries



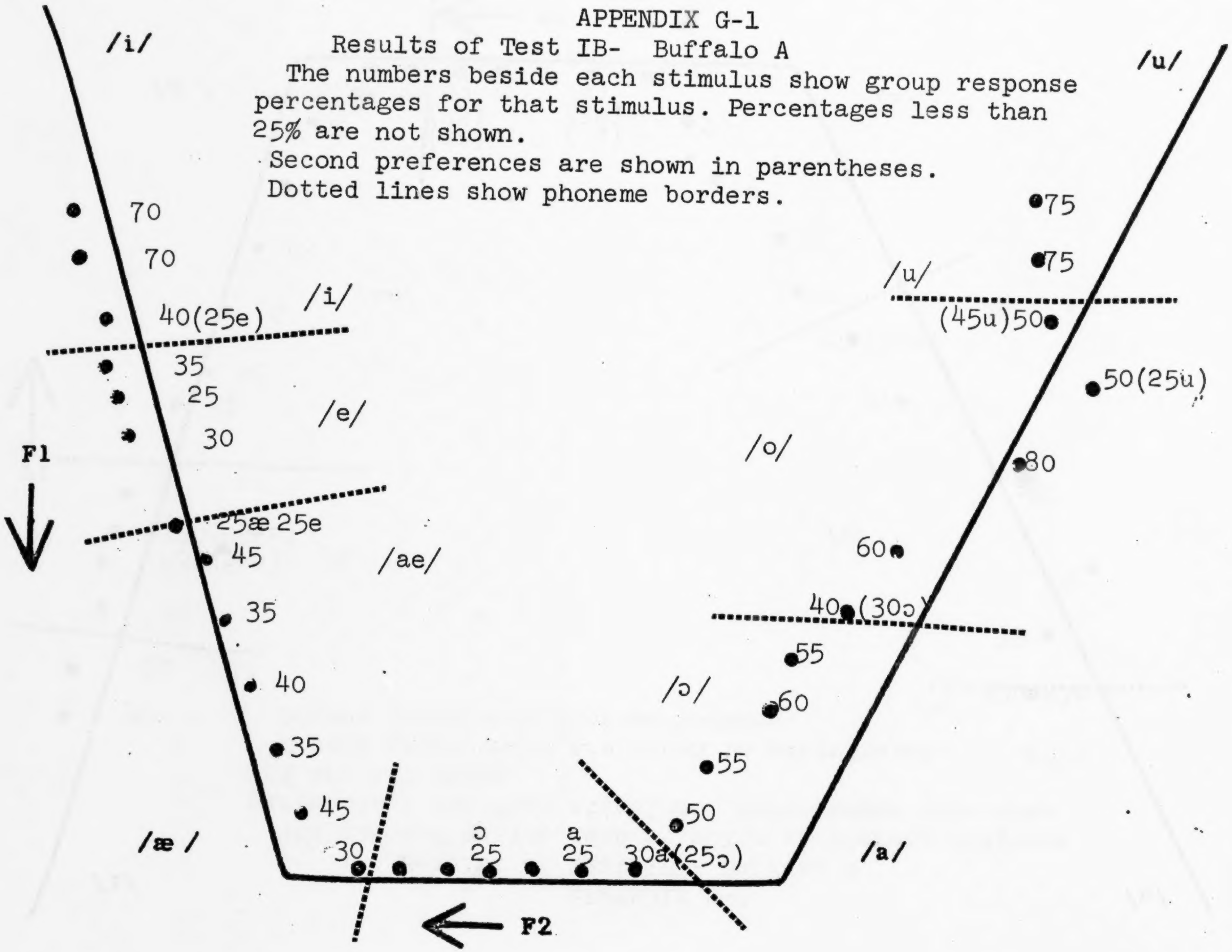
APPENDIX F-5

Results of Test IA - Virginia B

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.
Second preferences are shown in parentheses.
Dotted lines show phoneme boundaries.



APPENDIX G-1



APPENDIX G-1

Results of Test IB- Buffalo A

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.

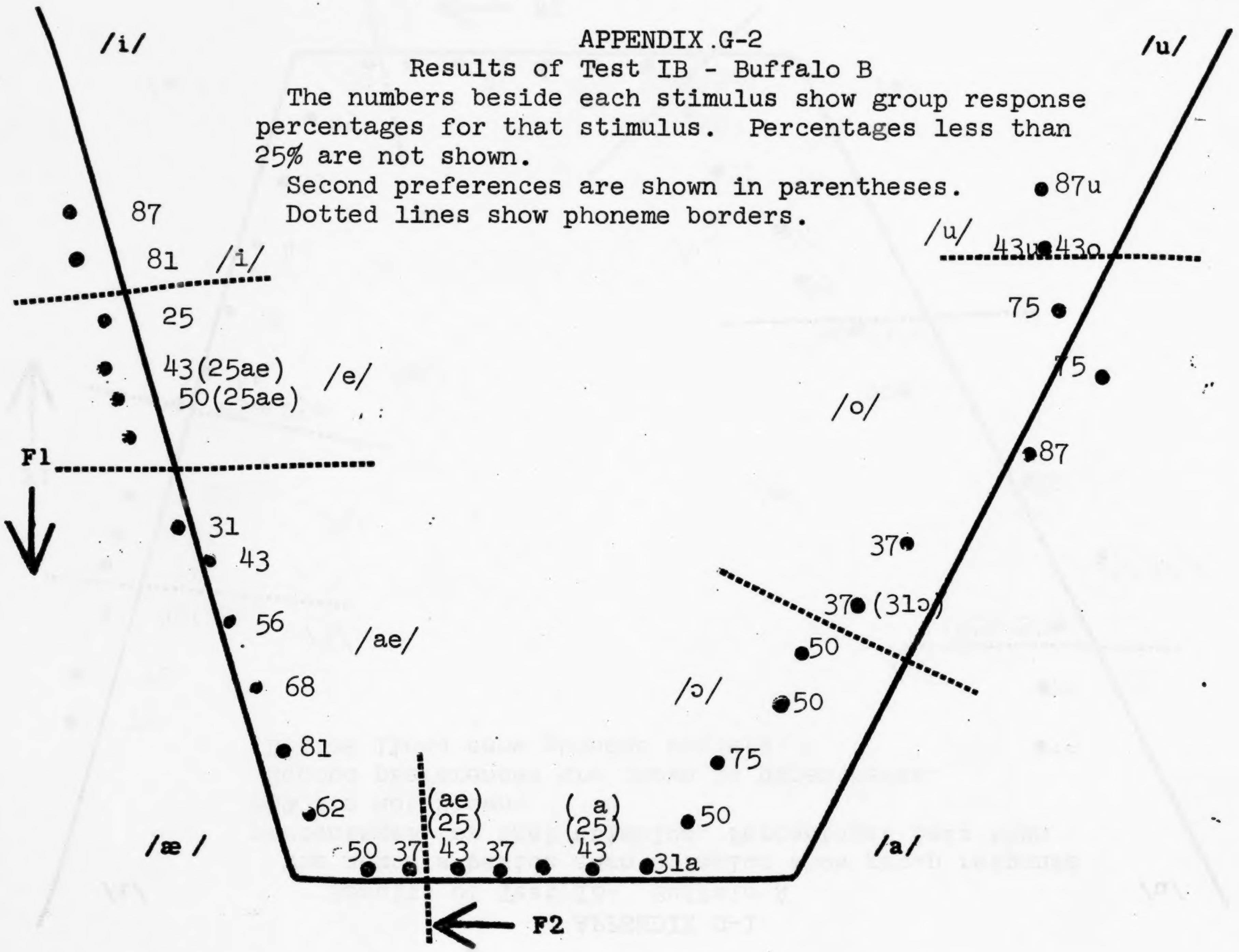
APPENDIX G-2

Results of Test IB - Buffalo B

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.



APPENDIX G-3

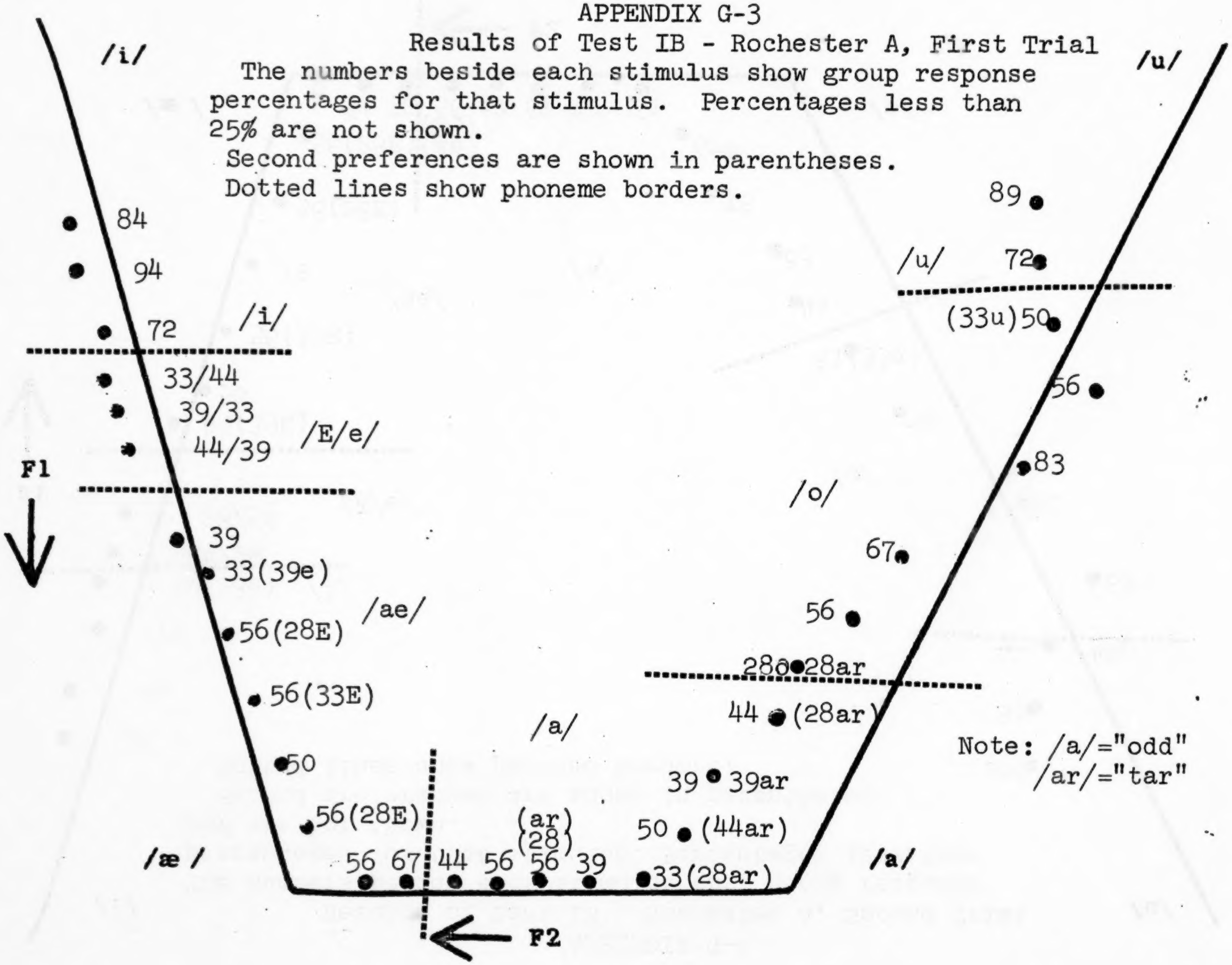
APPENDIX G-3

Results of Test IB - Rochester A, First Trial

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

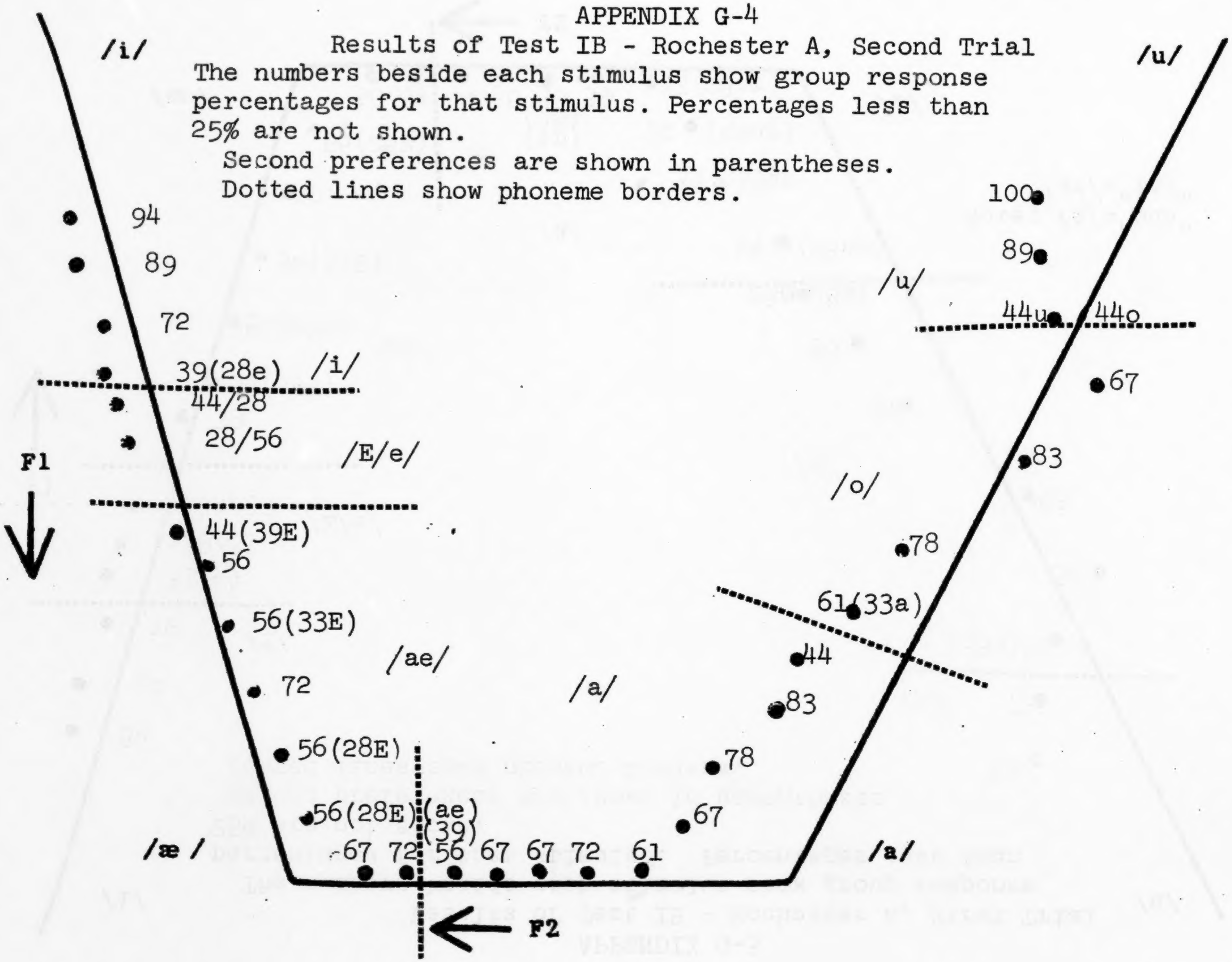
Second preferences are shown in parentheses.

Dotted lines show phoneme borders.



Note: /a/="odd"
/ar/="tar"

APPENDIX G-4



APPENDIX G-4

Results of Test IB - Rochester A, Second Trial

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.

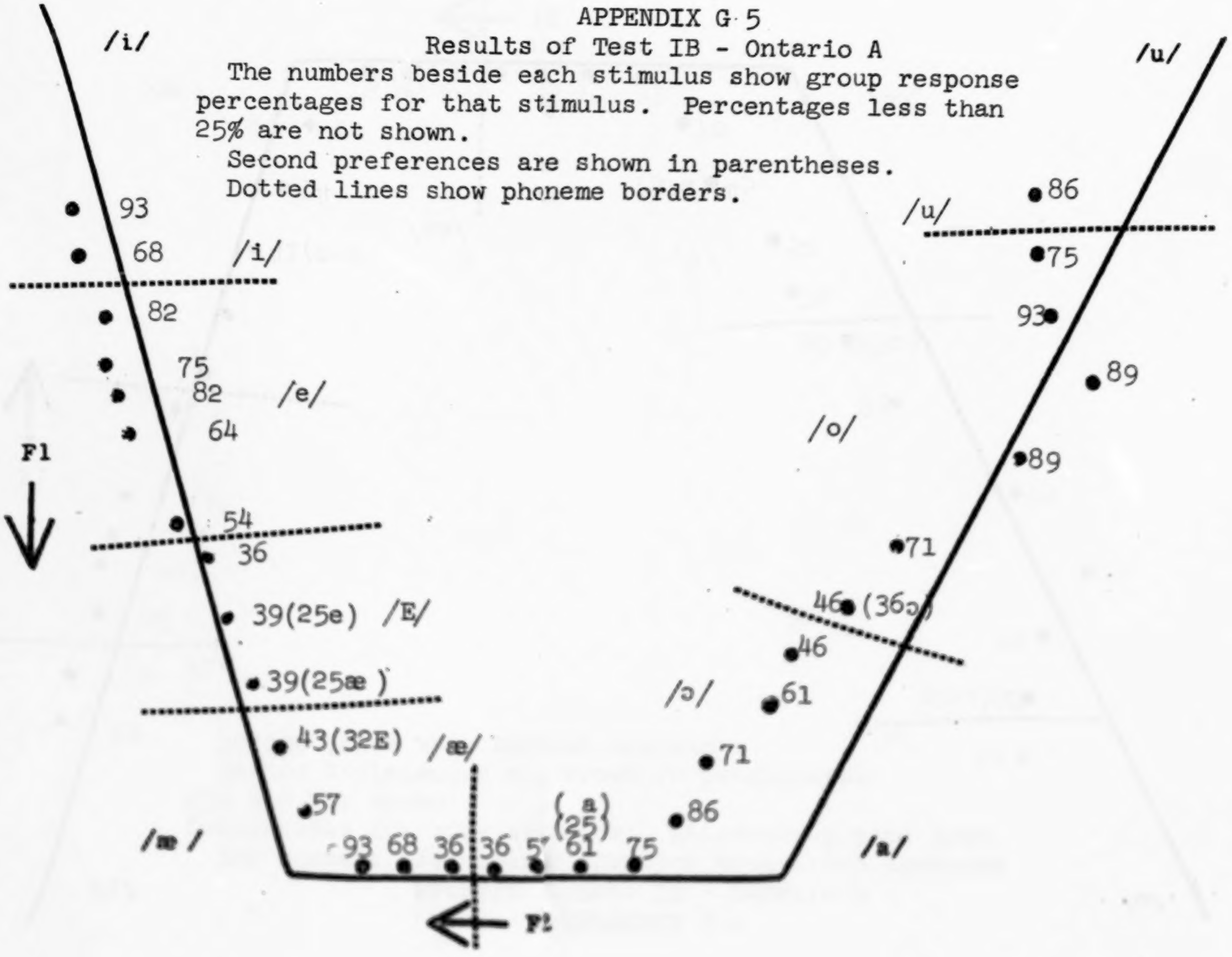
APPENDIX G 5

Results of Test IB - Ontario A

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.



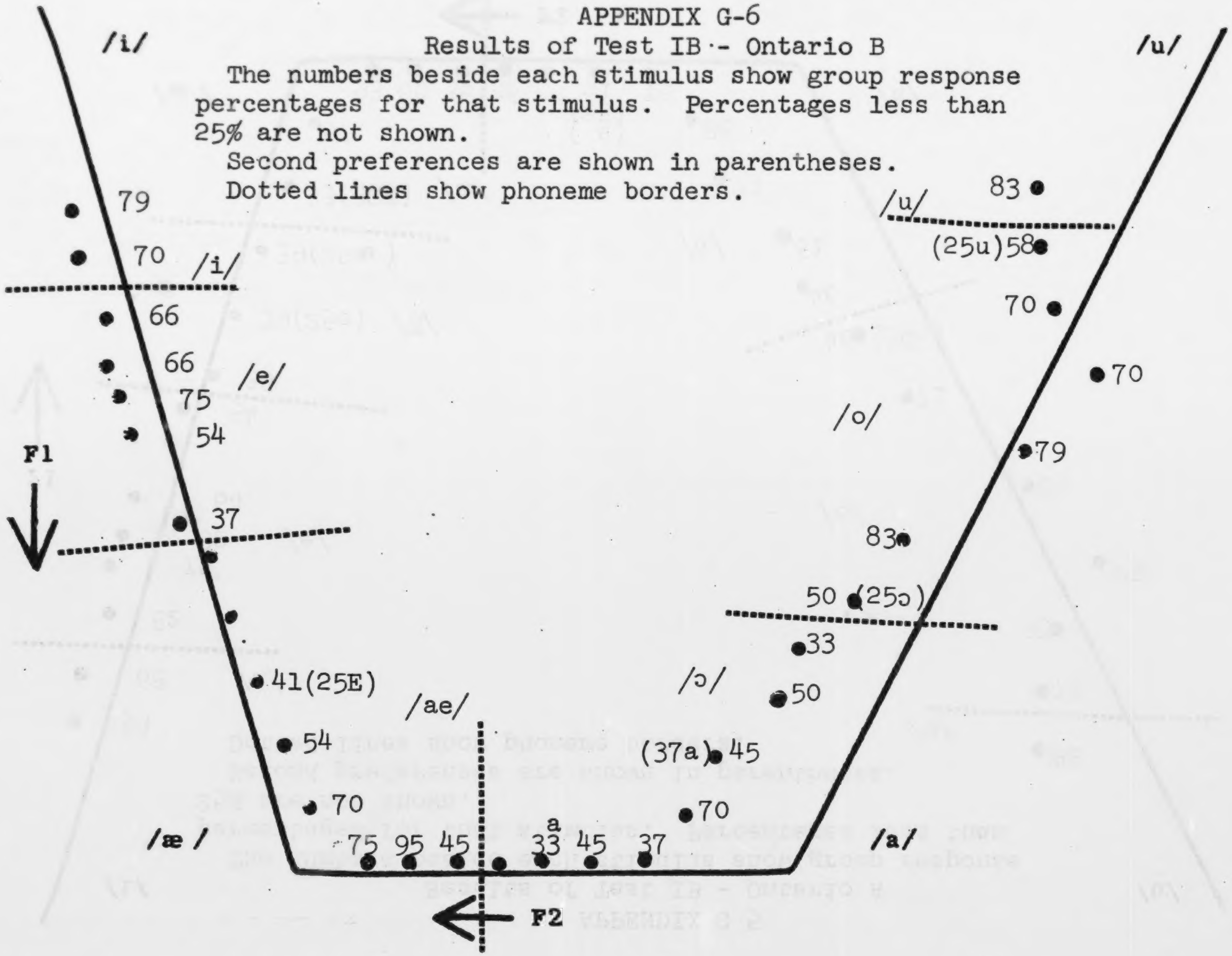
APPENDIX G-6

Results of Test IB - Ontario B

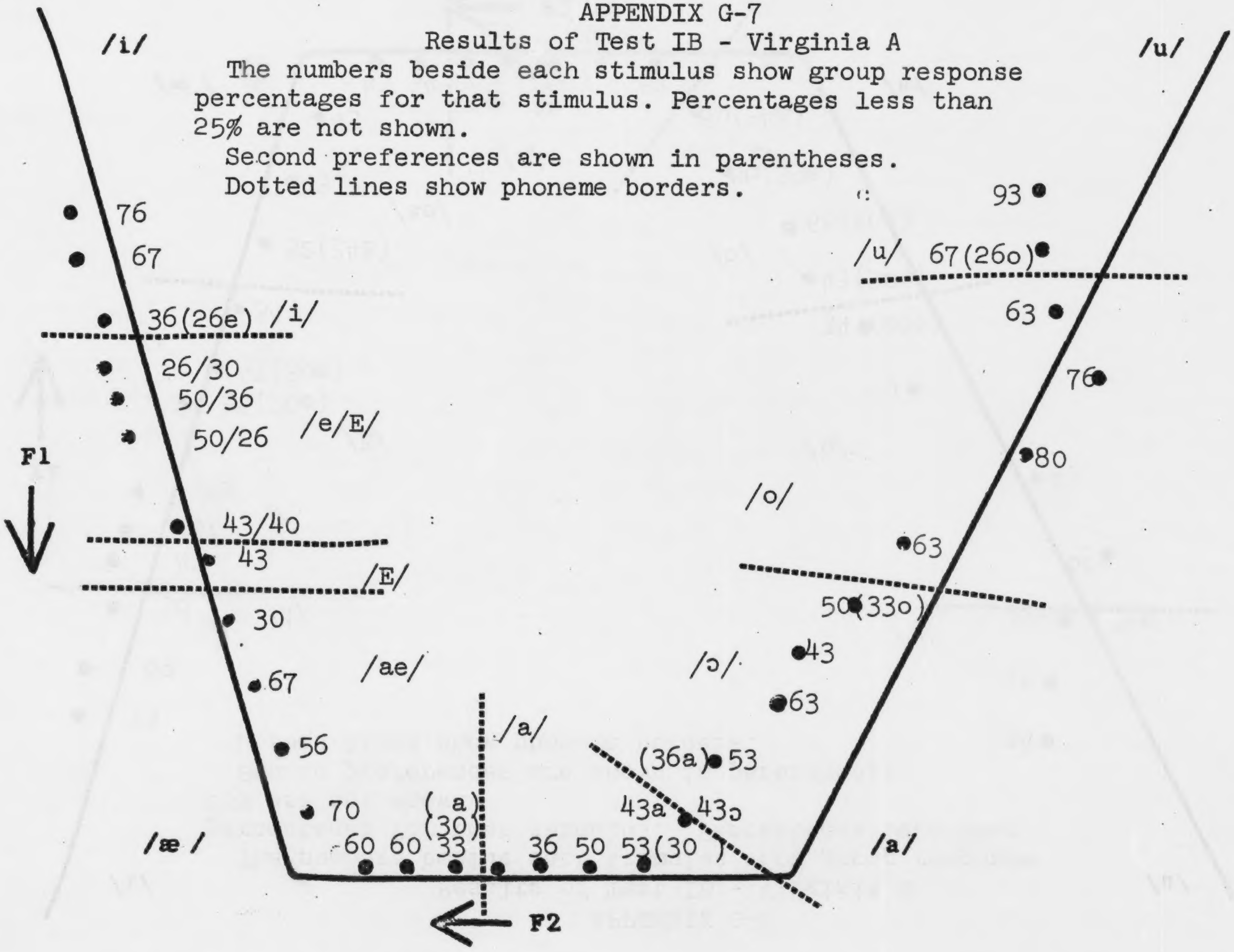
The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.



APPENDIX G-7



APPENDIX G-7

Results of Test IB - Virginia A

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.

APPENDIX G-8

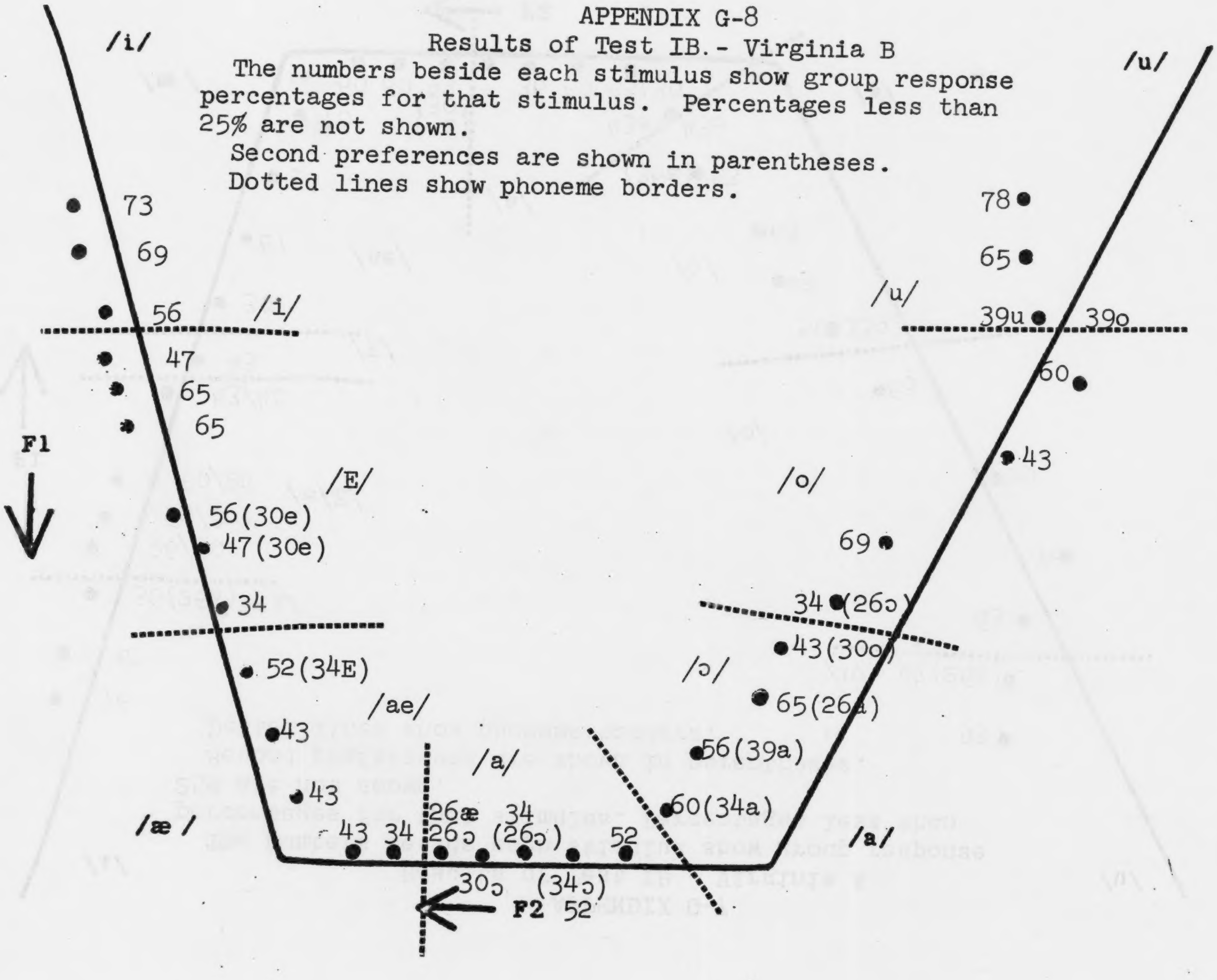
APPENDIX G-8

Results of Test IB.- Virginia B

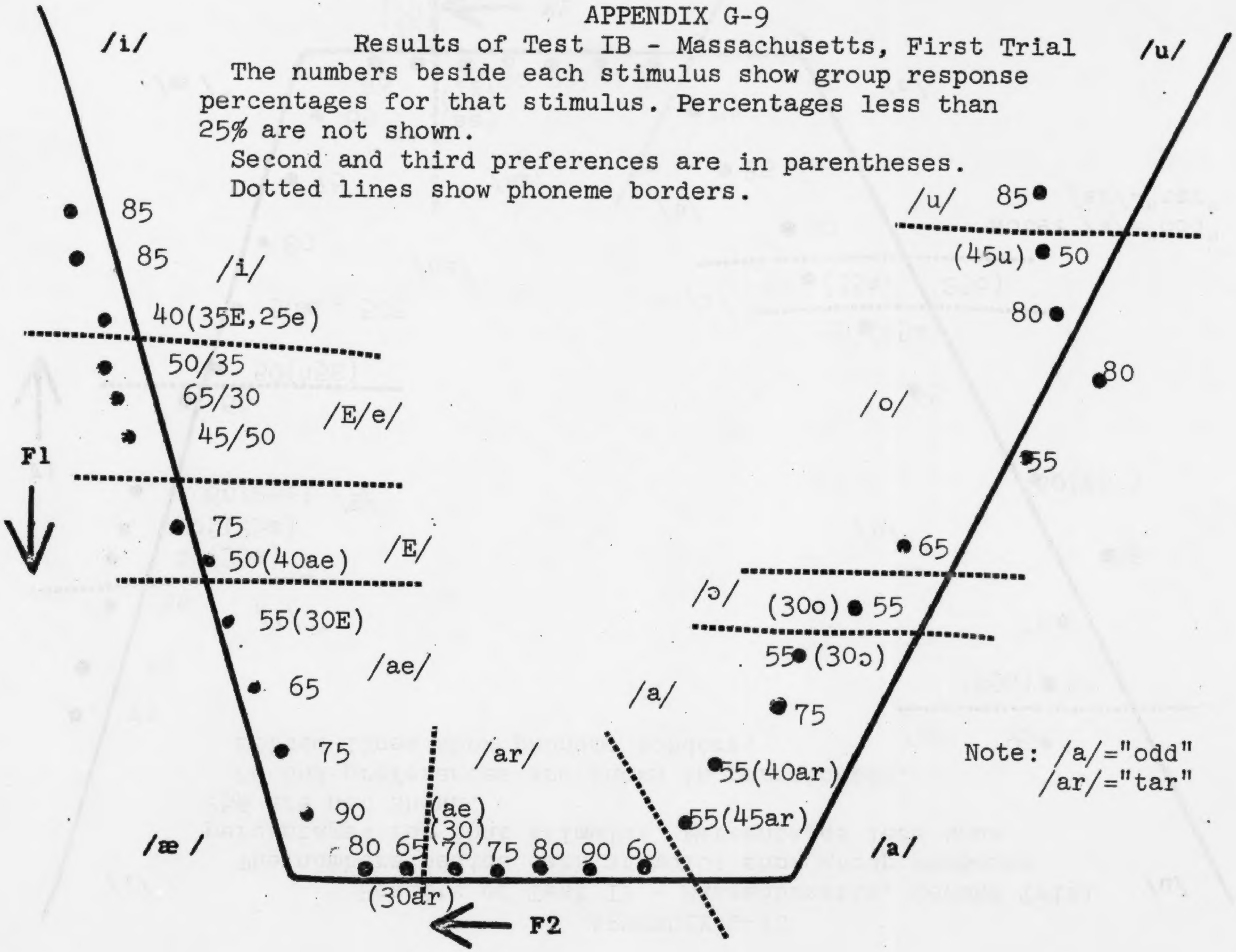
The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.



APPENDIX G-9



APPENDIX G-9

Results of Test IB - Massachusetts, First Trial

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second and third preferences are in parentheses.

Dotted lines show phoneme borders.

Note: /a/="odd"
/ar/="tar"

APPENDIX G-10

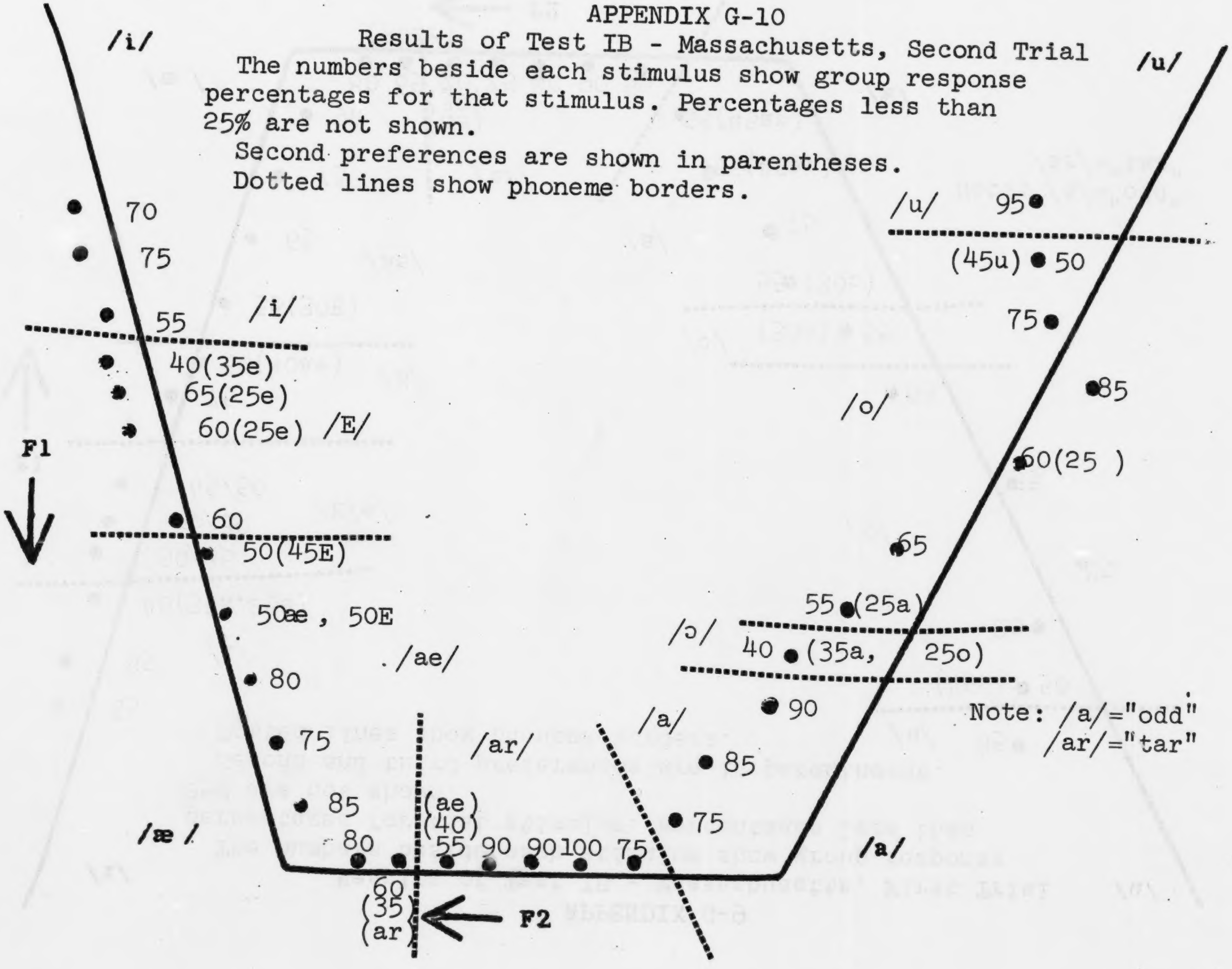
APPENDIX G-10

Results of Test IB - Massachusetts, Second Trial

The numbers beside each stimulus show group response percentages for that stimulus. Percentages less than 25% are not shown.

Second preferences are shown in parentheses.

Dotted lines show phoneme borders.



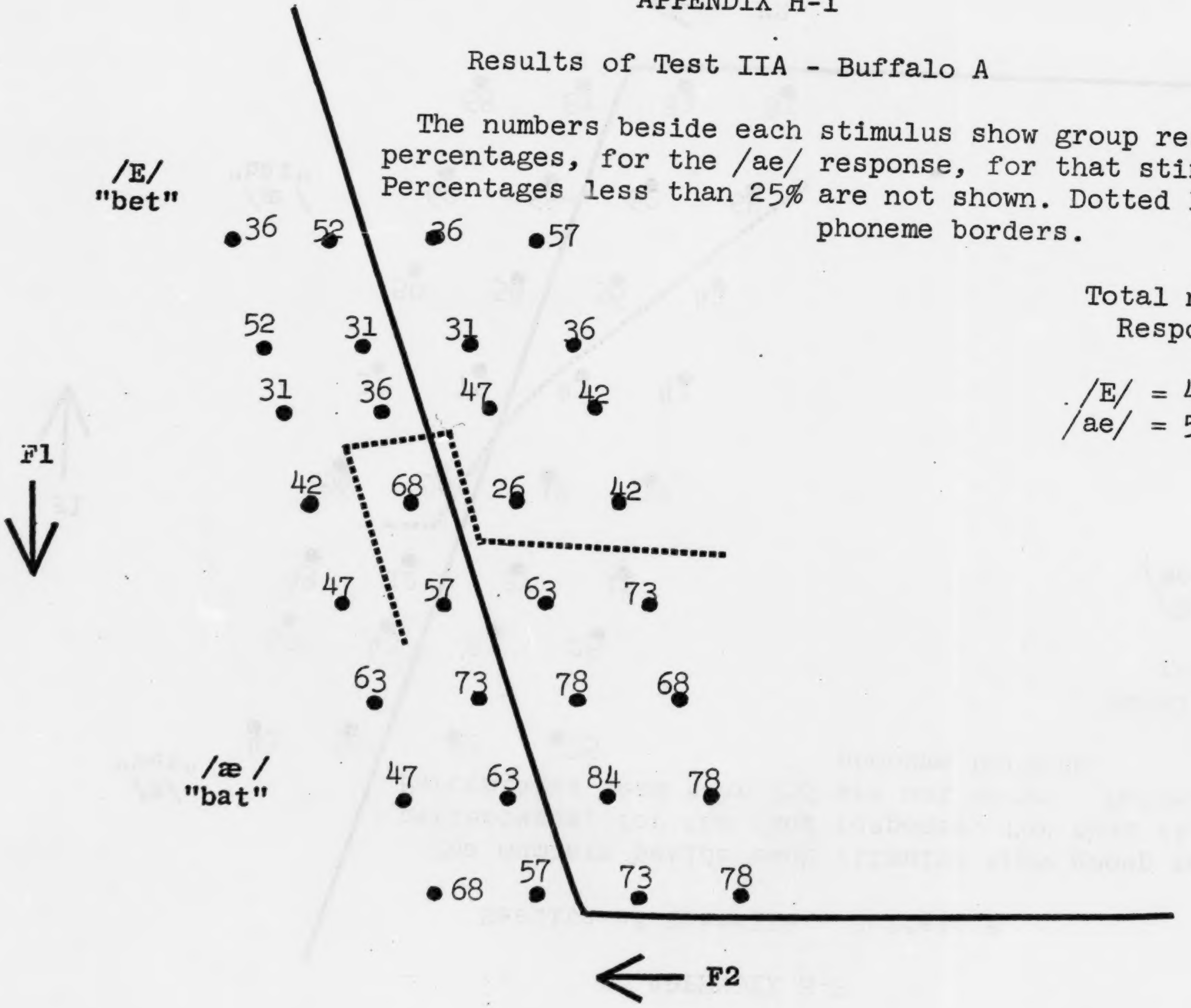
APPENDIX H-1

Results of Test IIA - Buffalo A

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted lines show phoneme borders.

Total number of Responses

/E/ = 42%
/ae/ = 54%



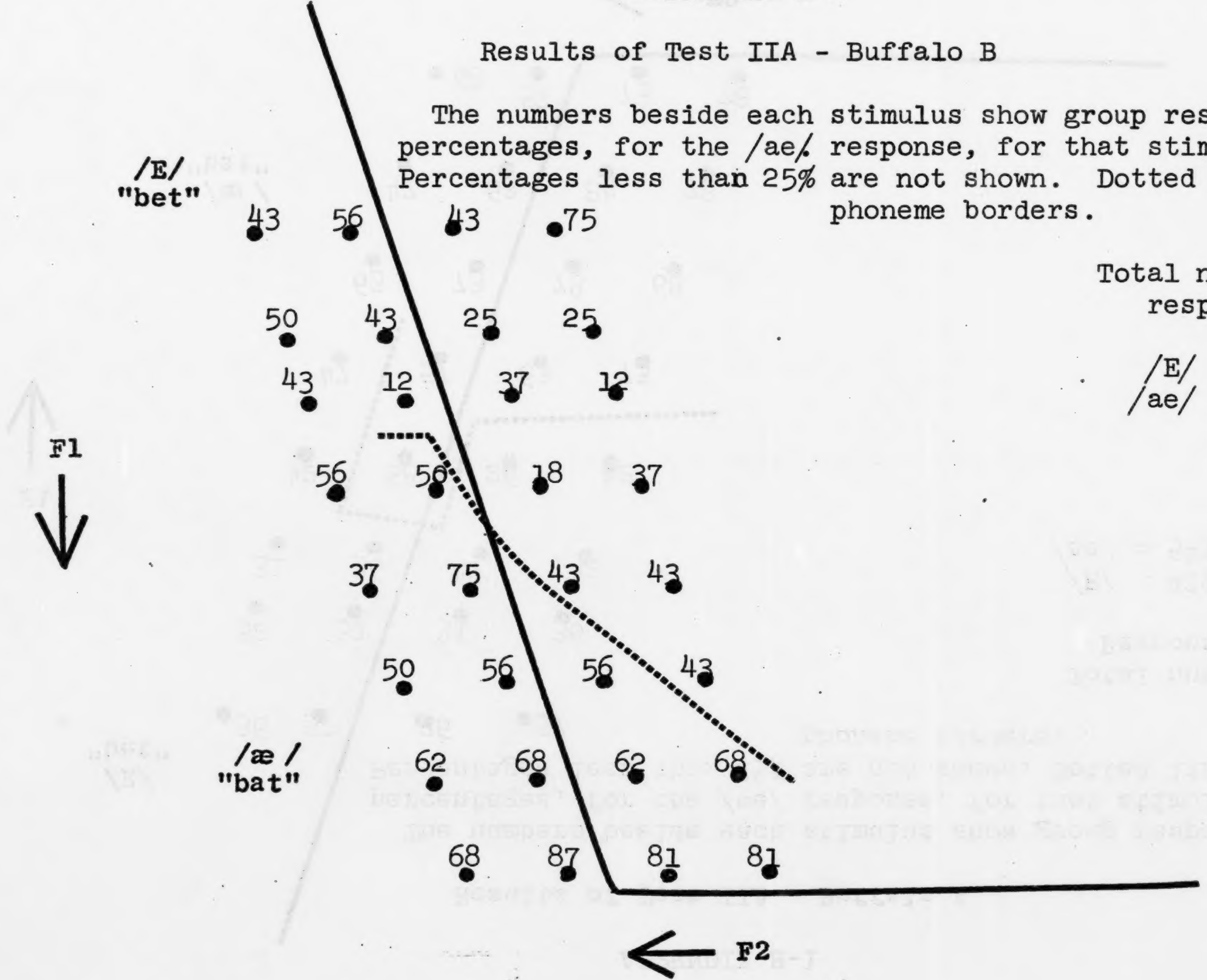
APPENDIX H-2

Results of Test IIA - Buffalo B

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted lines show phoneme borders.

Total number of responses

/E/ = 48%
/ae/ = 51%



APPENDIX H-3

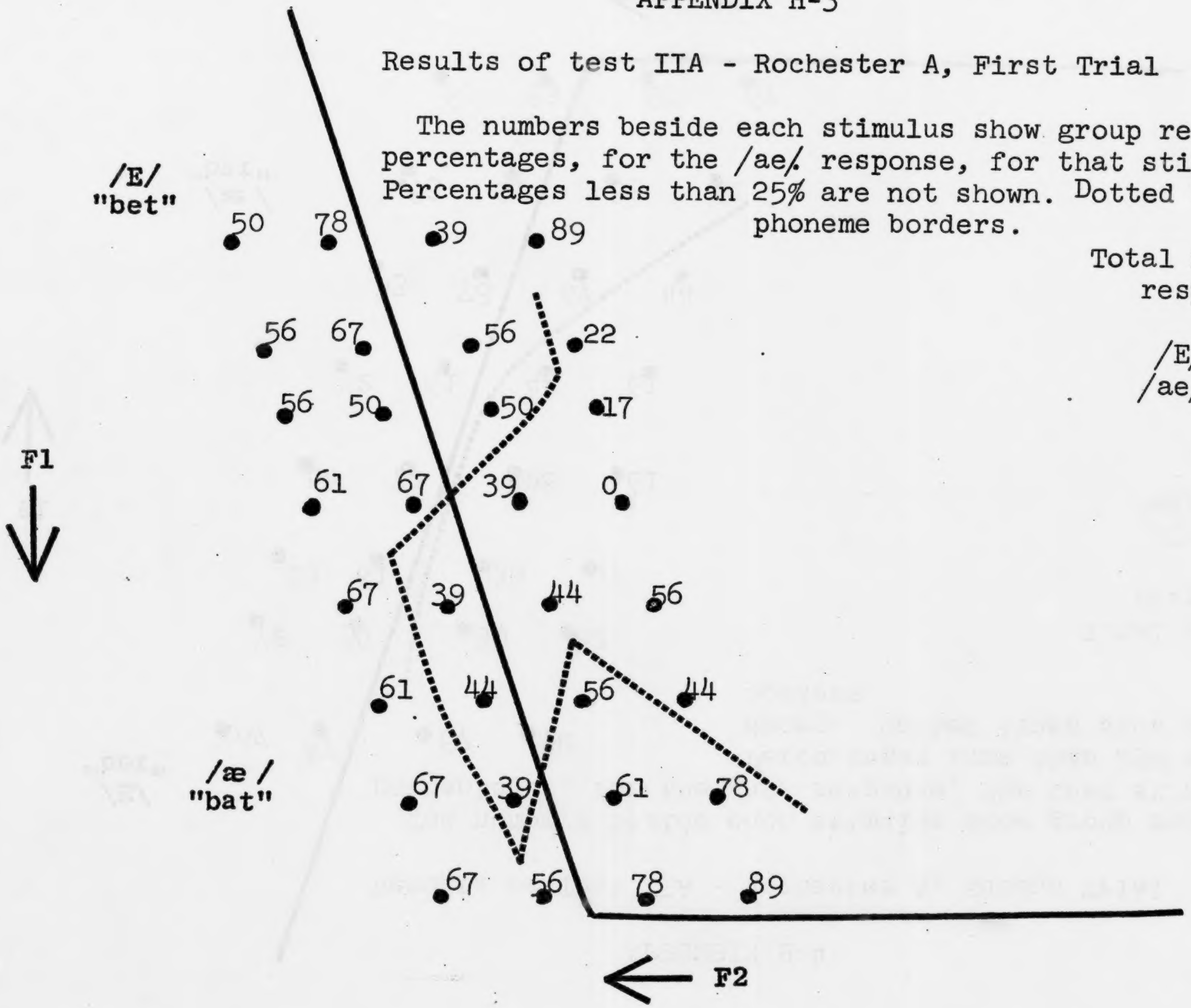
APPENDIX H-3

Results of test IIA - Rochester A, First Trial

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted lines show phoneme borders.

Total number of responses

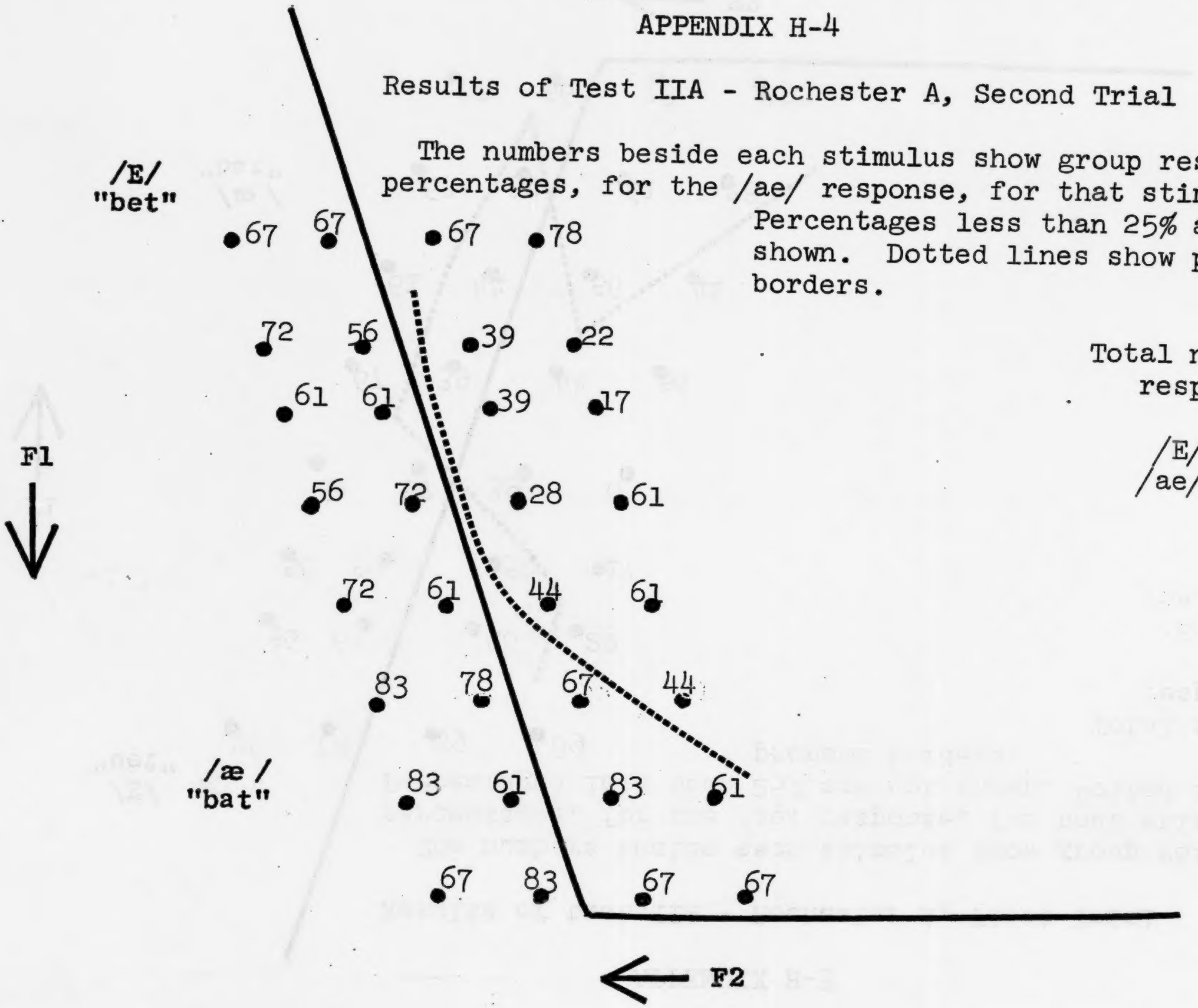
/E/ = 46%
/ae/ = 54%



APPENDIX H-4

Results of Test IIA - Rochester A, Second Trial

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted lines show phoneme borders.



Total number of responses

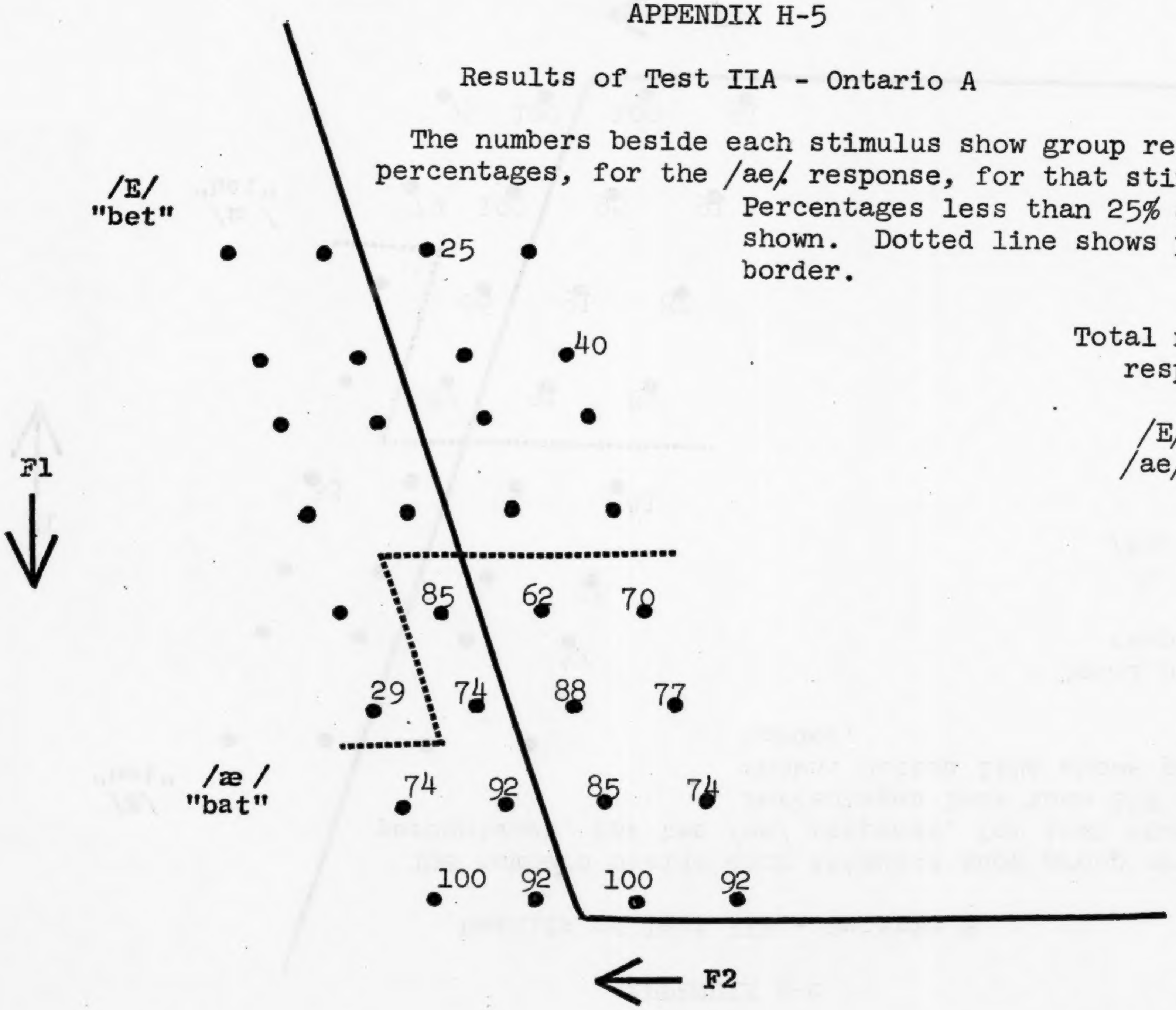
/E/ = 48%
/ae/ = 52%

APPENDIX H-5

APPENDIX H-5

Results of Test IIA - Ontario A

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted line shows phoneme border.



Total number of responses

/E/ = 53%
/ae/ = 47%

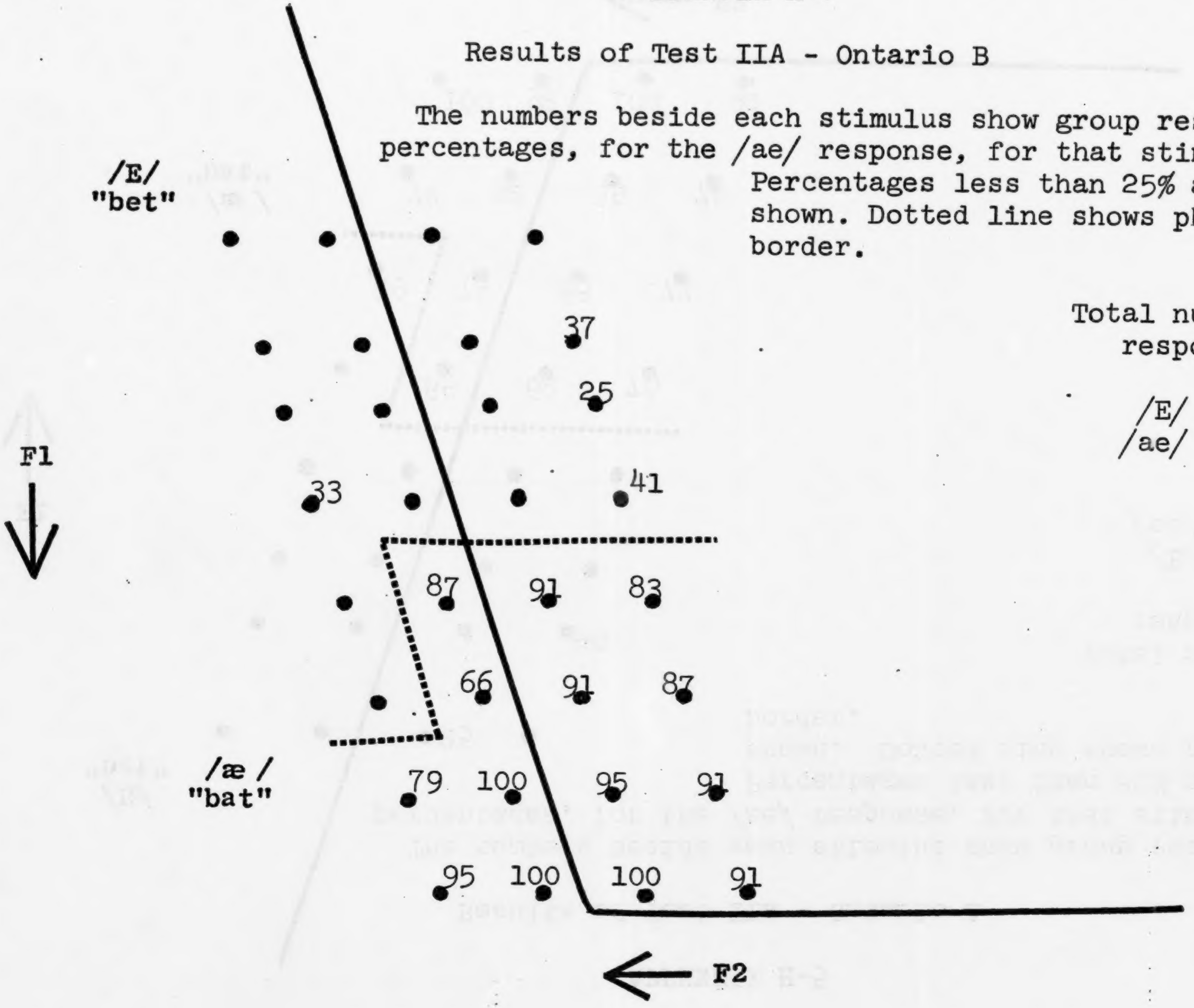
APPENDIX H-6

Results of Test IIA - Ontario B

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted line shows phoneme border.

Total number of responses

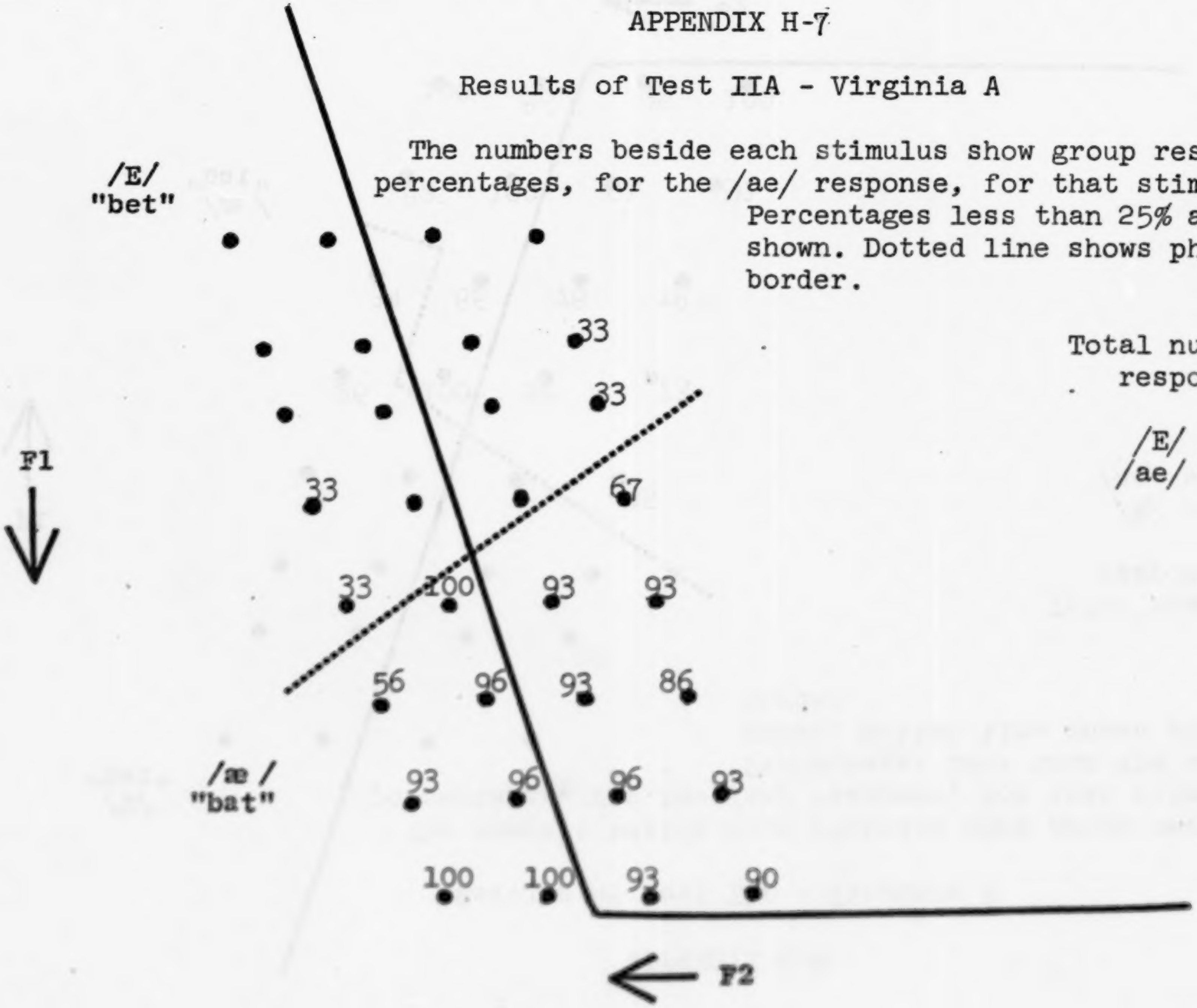
/E/ = 51%
/ae/ = 49%



APPENDIX H-7

Results of Test IIA - Virginia A

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted line shows phoneme border.



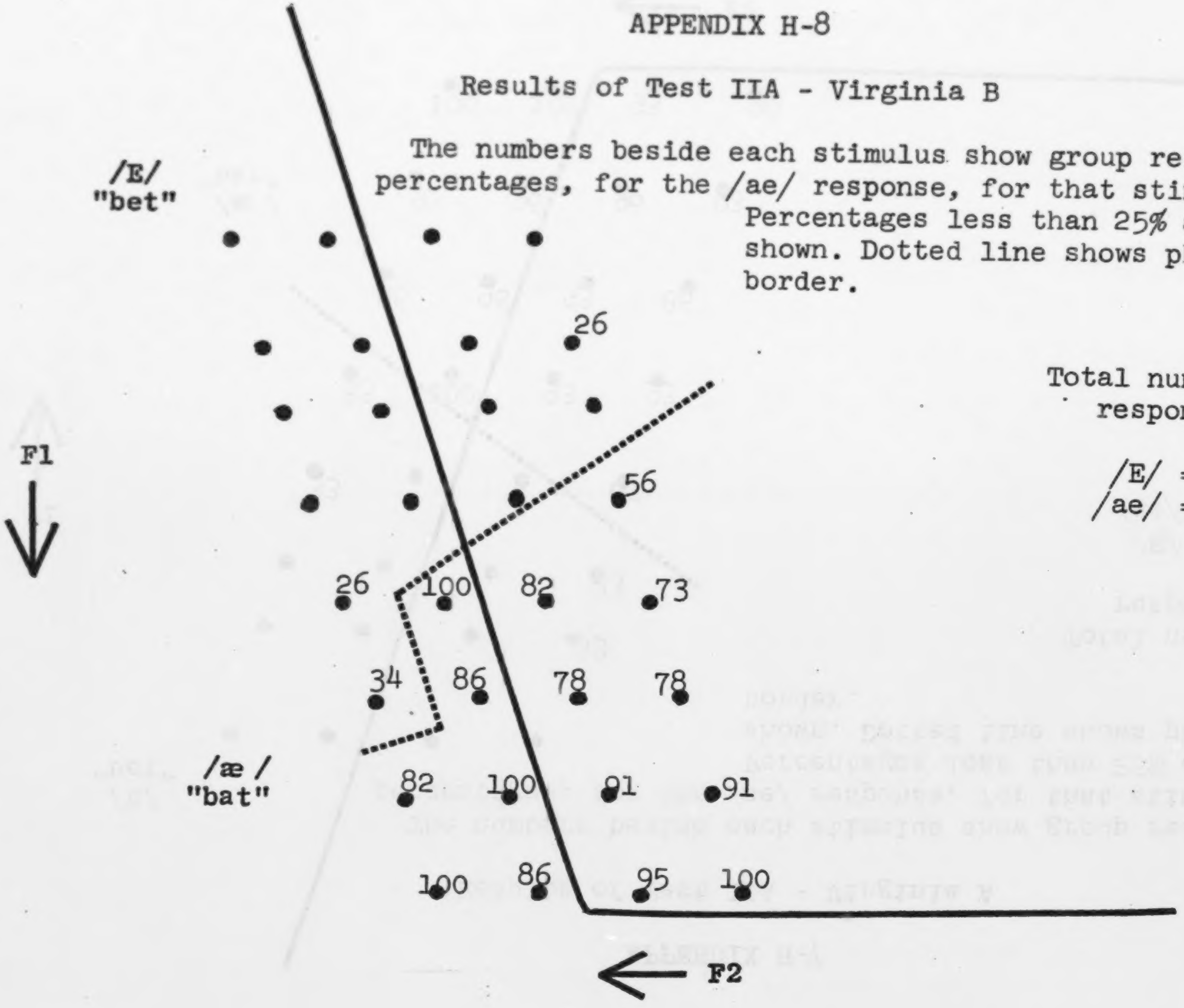
Total number of responses

/E/ = 47%
/ae/ = 53%

APPENDIX H-8

Results of Test IIA - Virginia B

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted line shows phoneme border.



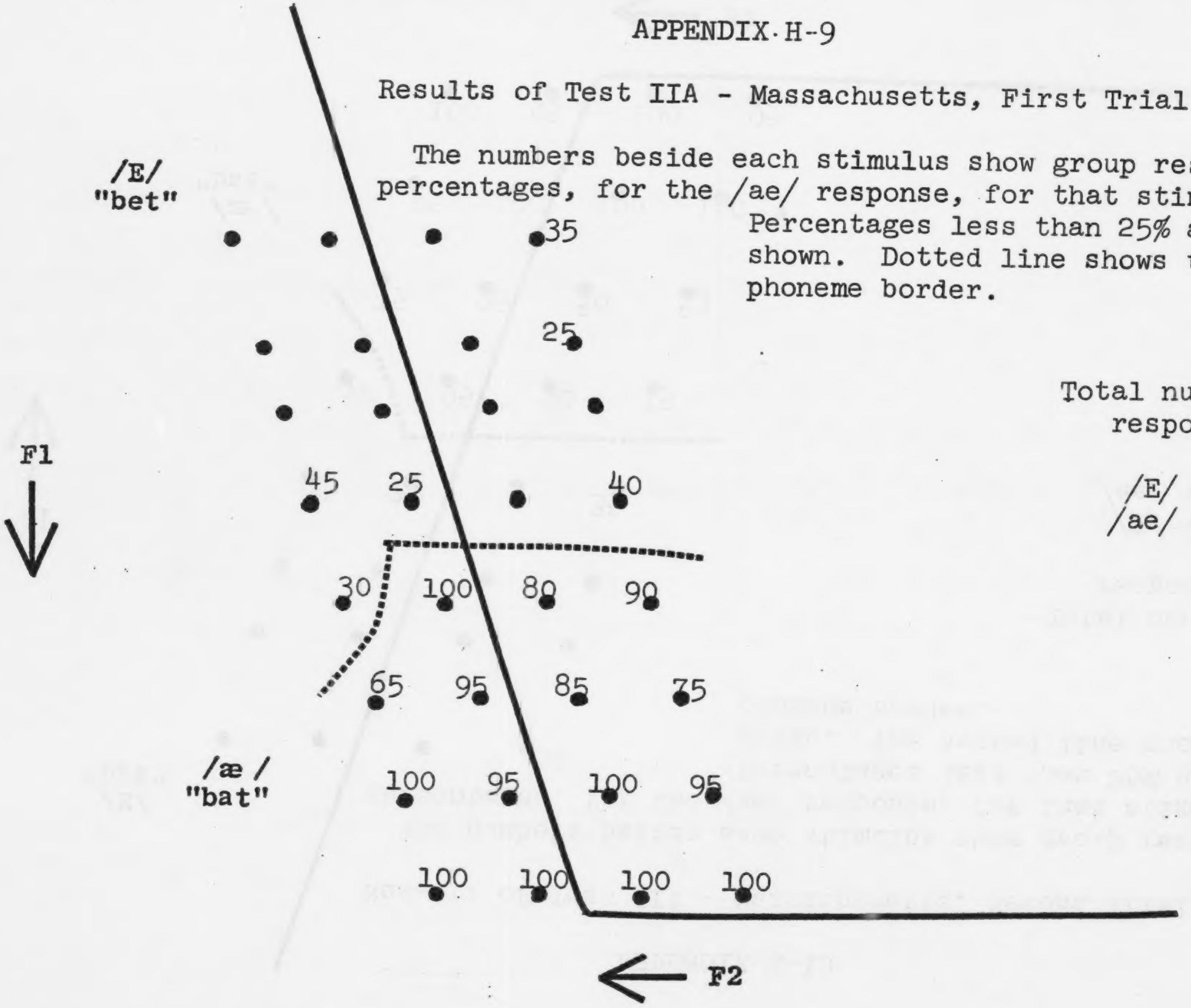
Total number of responses

/E/ = 52%
/ae/ = 47%

APPENDIX H-9

Results of Test IIA - Massachusetts, First Trial

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. Dotted line shows the phoneme border.



Total number of responses

/E/ = 47%
/ae/ = 53%

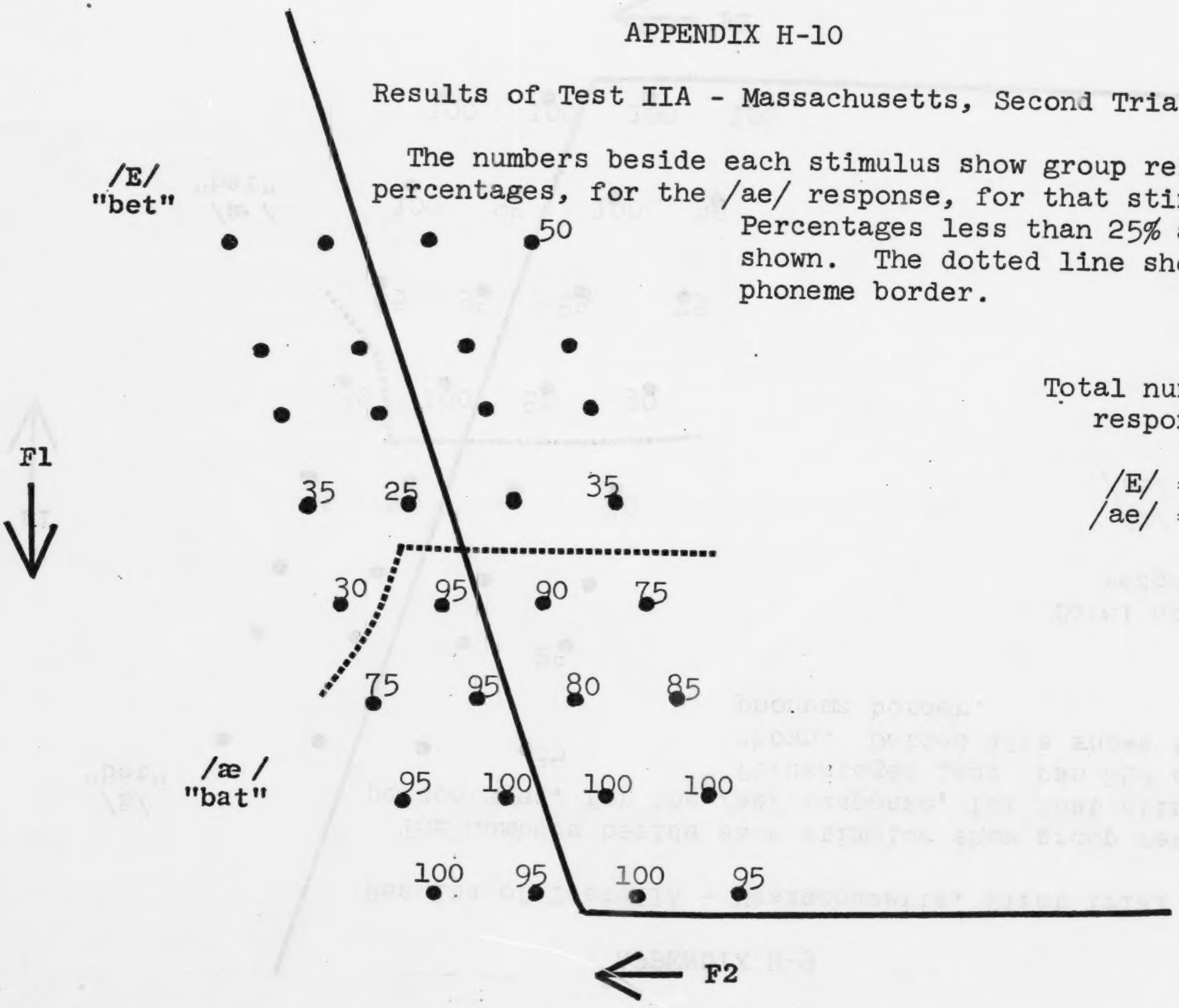
APPENDIX H-10

Results of Test IIA - Massachusetts, Second Trial

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus. Percentages less than 25% are not shown. The dotted line shows the phoneme border.

Total number of responses

/E/ = 47%
/ae/ = 52%



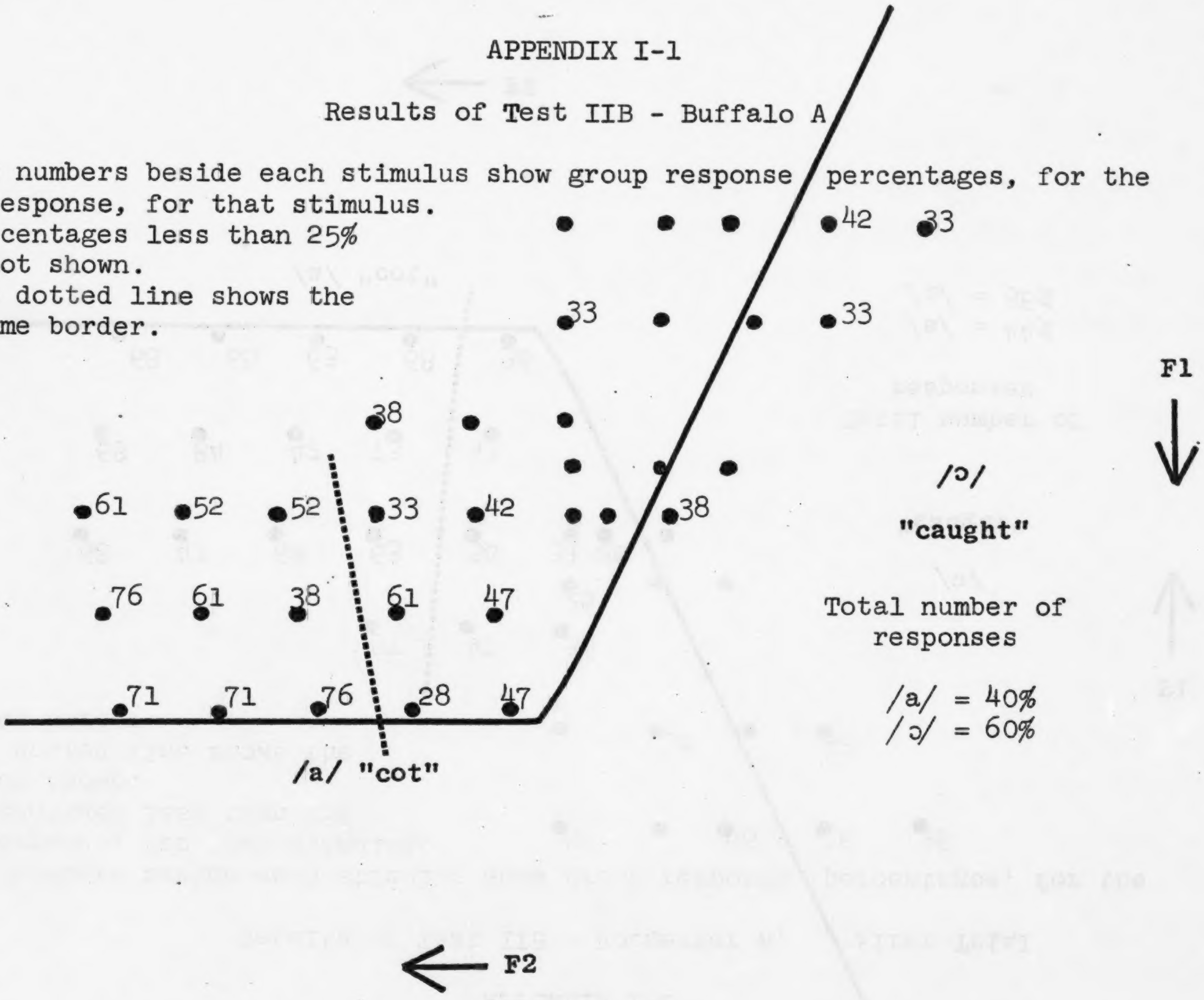
APPENDIX I-1

Results of Test IIB - Buffalo A

The numbers beside each stimulus show group response percentages, for the /a/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.



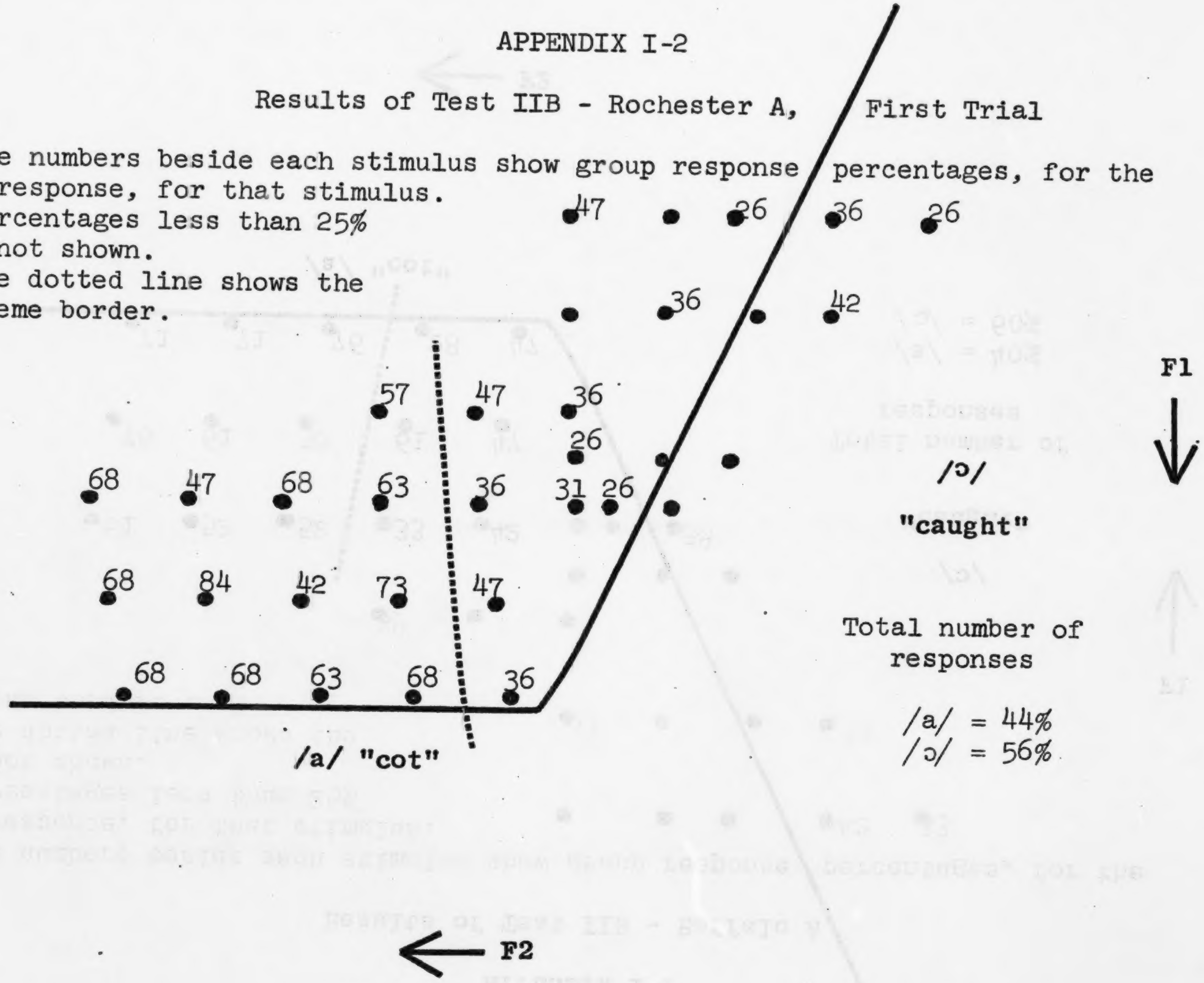
APPENDIX I-2

Results of Test IIB - Rochester A, First Trial

The numbers beside each stimulus show group response percentages, for the /a/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.

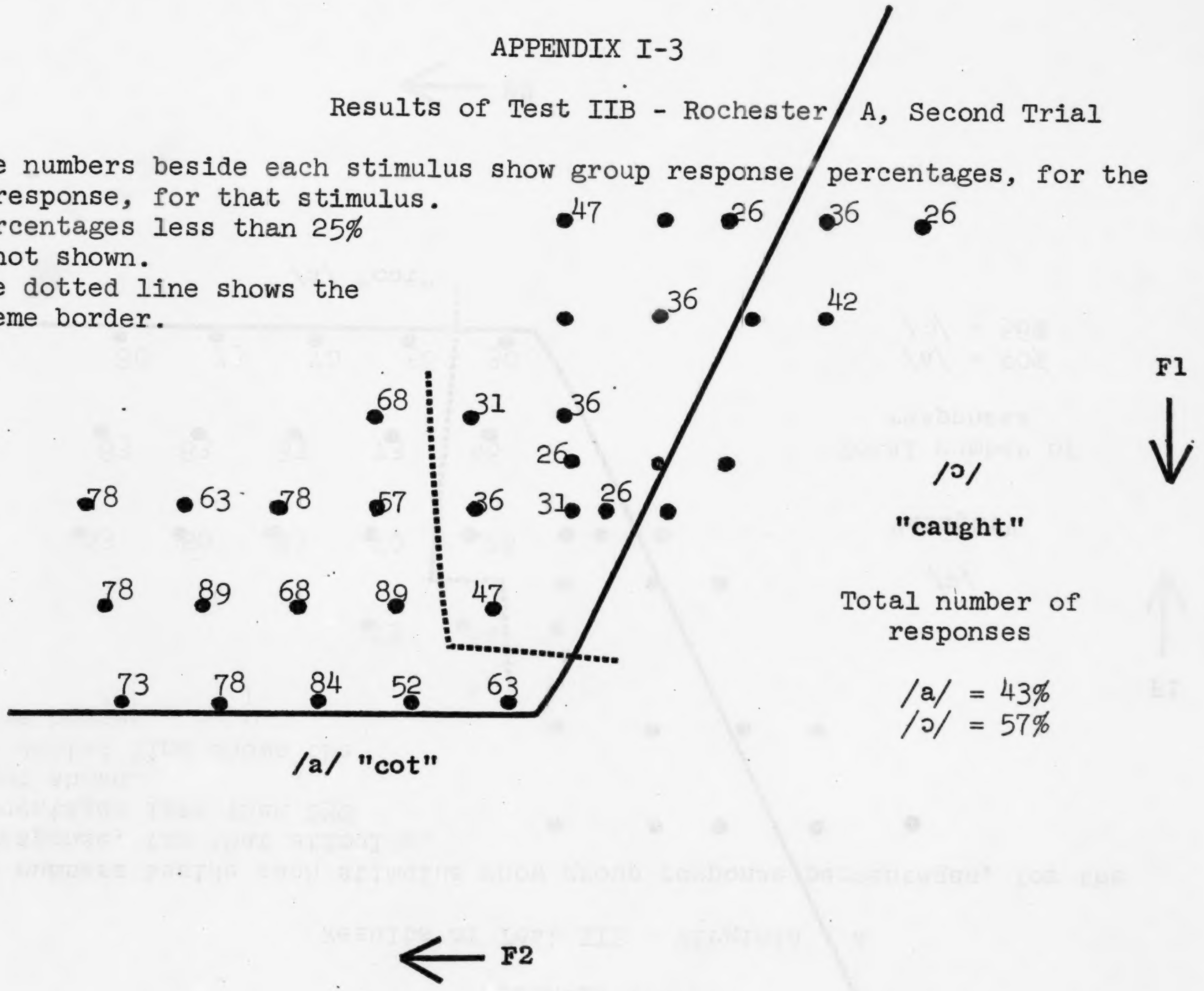


Total number of responses

/a/ = 44%
/ɔ/ = 56%

APPENDIX I-3

Results of Test IIB - Rochester A, Second Trial



The numbers beside each stimulus show group response percentages, for the /a/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.

Total number of responses

/a/ = 43%
/ɔ/ = 57%

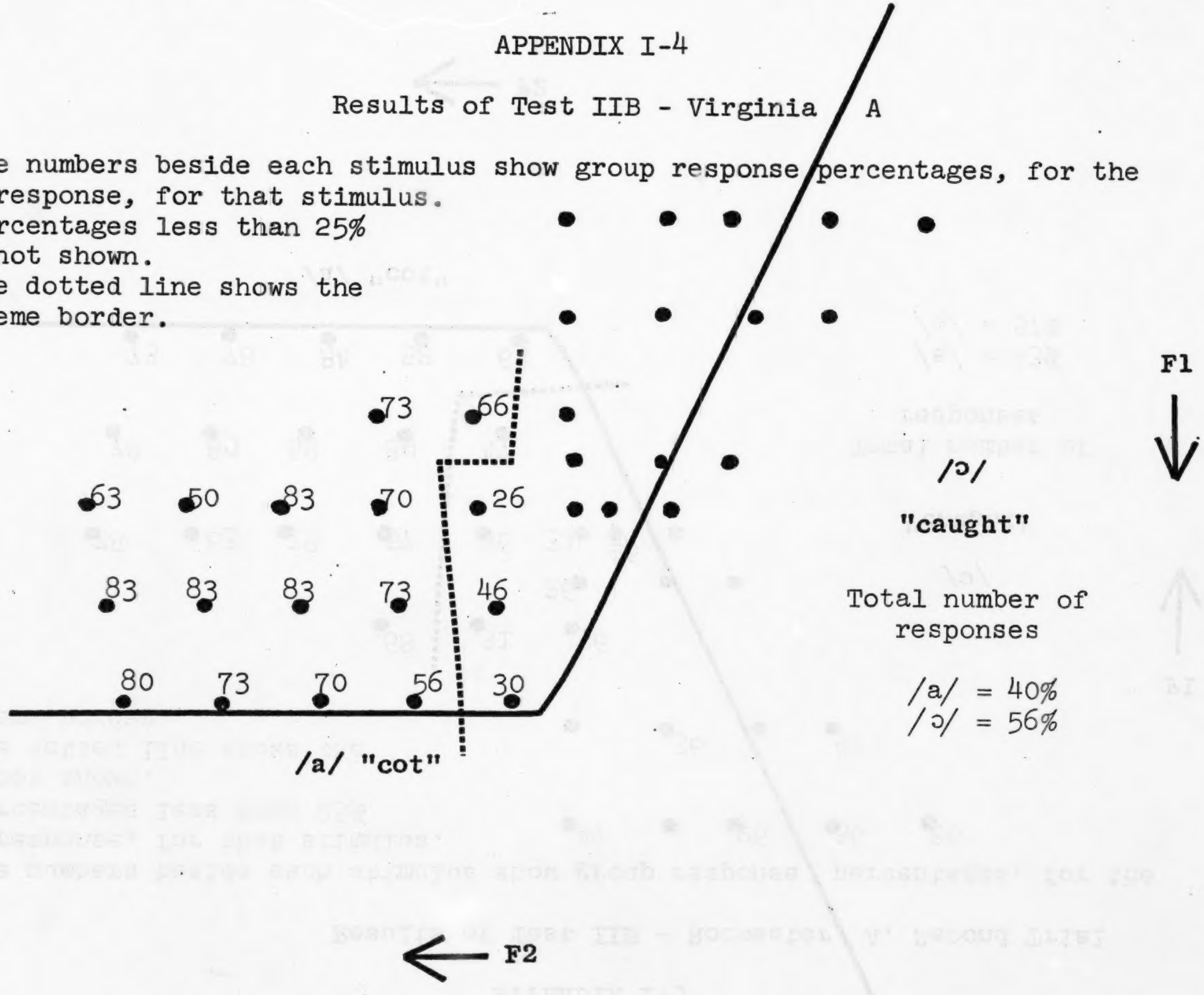
APPENDIX I-4

Results of Test IIB - Virginia A

The numbers beside each stimulus show group response percentages, for the /a/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.



Total number of responses

/a/ = 40%
/ɔ/ = 56%

APPENDIX I-5

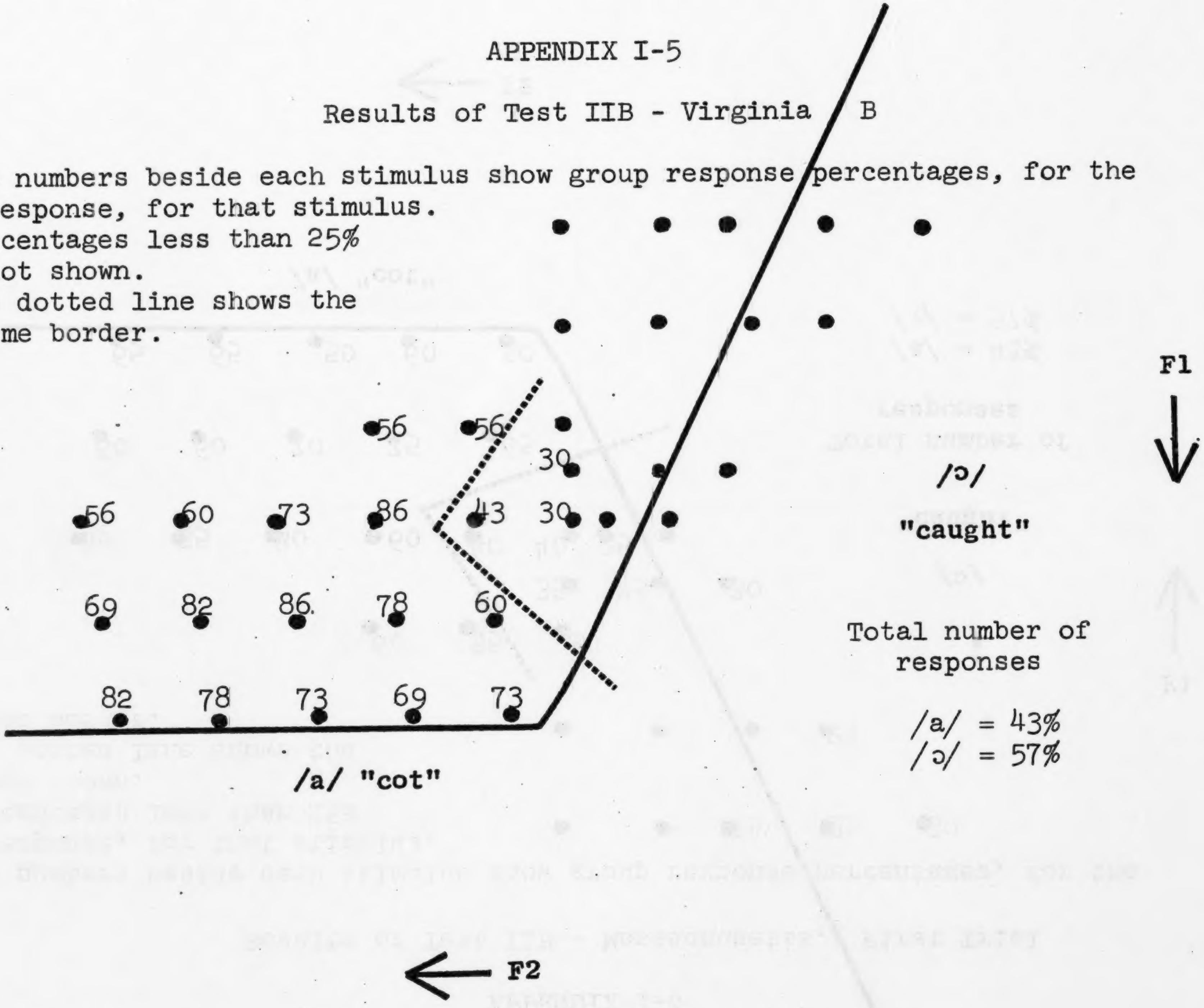
APPENDIX I-5

Results of Test IIB - Virginia B

The numbers beside each stimulus show group response percentages, for the /a/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.



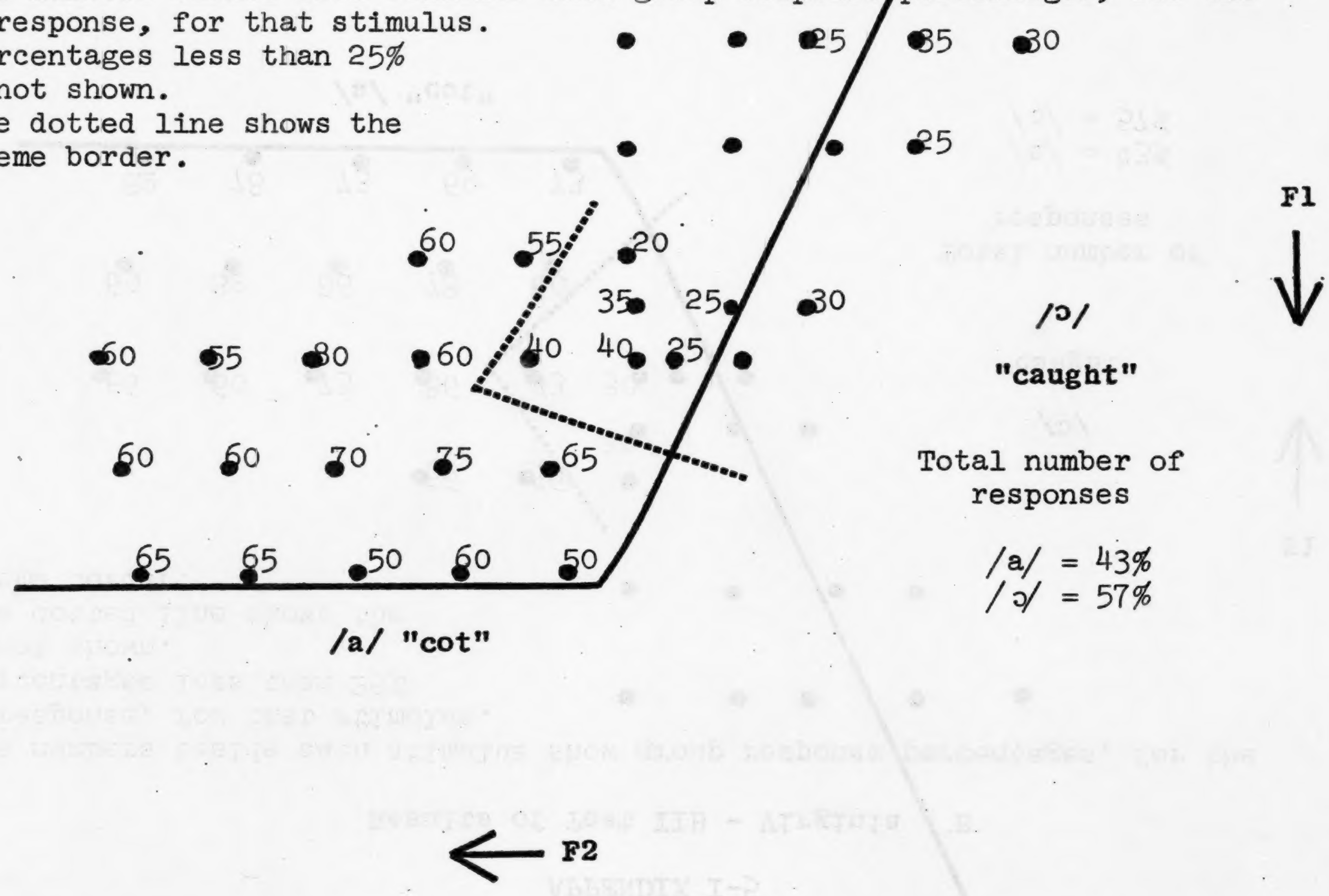
APPENDIX I-6

Results of Test IIB - Massachusetts, First Trial

The numbers beside each stimulus show group response percentages, for the /a/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.



Total number of responses

/a/ = 43%
/ɔ/ = 57%

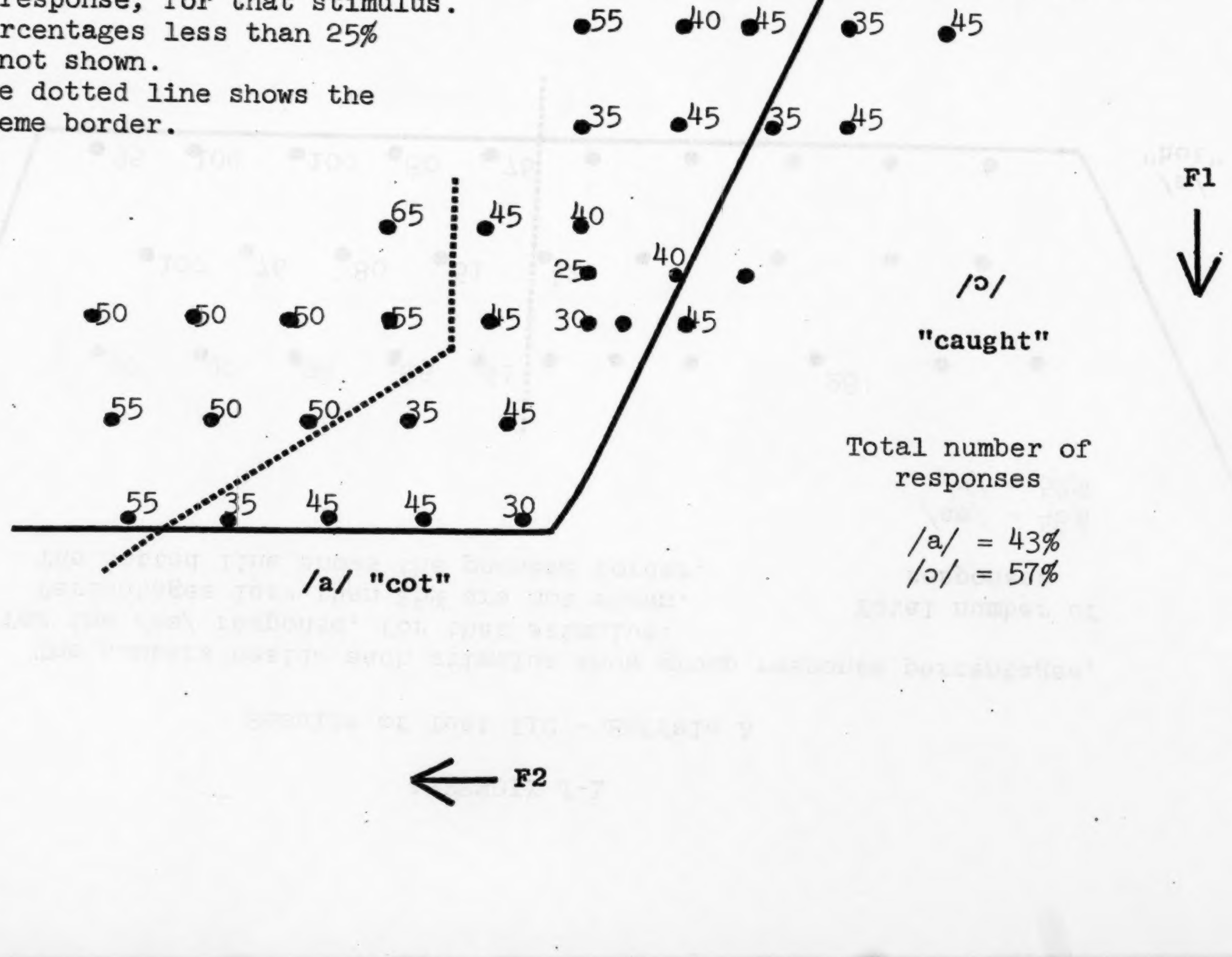
APPENDIX I-7

Results of Test IIB - Massachusetts, First Trial

The numbers beside each stimulus show group response percentages, for the /a/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.



Total number of responses

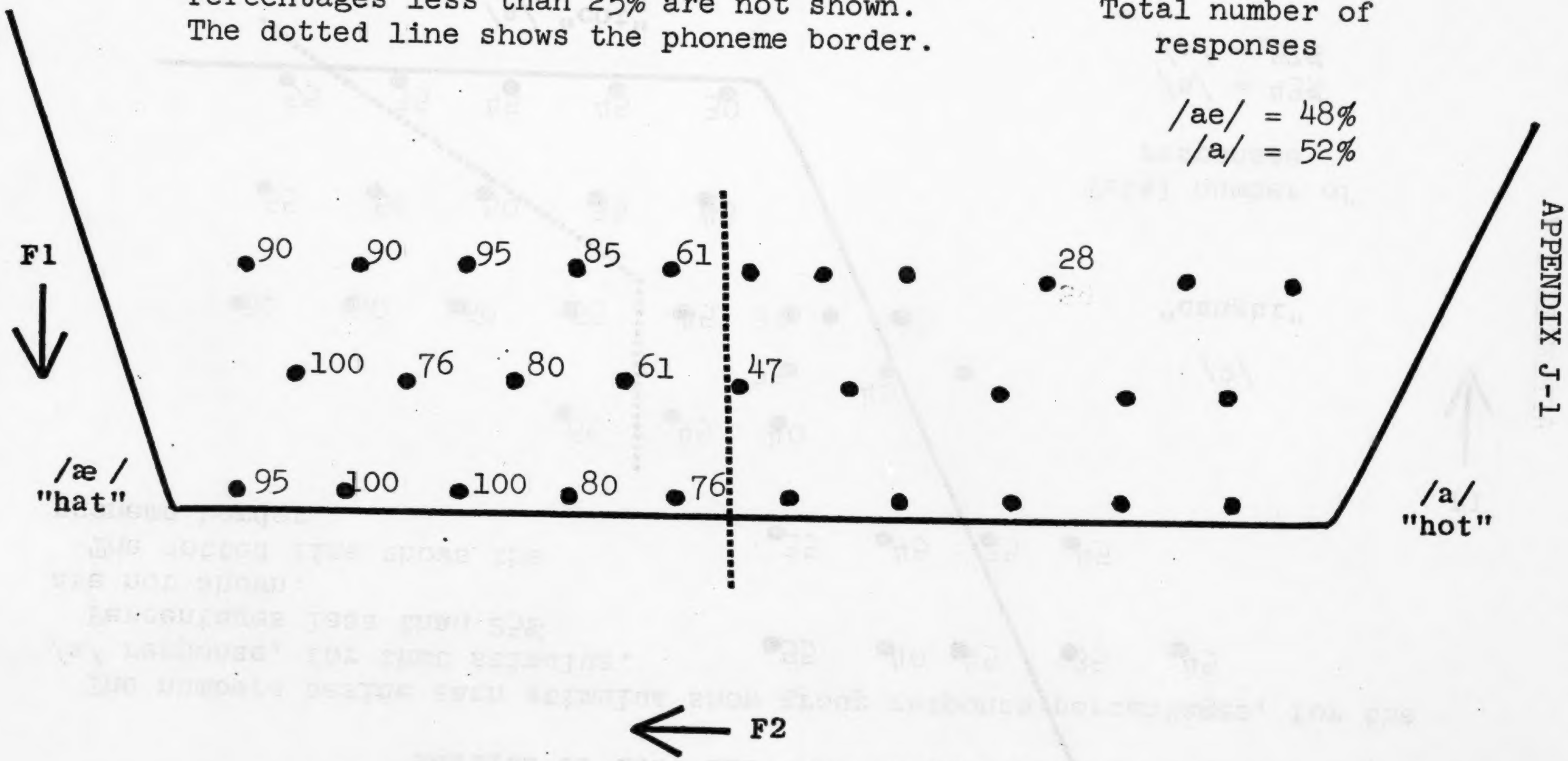
/a/ = 43%

/ɔ/ = 57%

APPENDIX J-1

Results of Test IIC - Buffalo A

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus.
Percentages less than 25% are not shown.
The dotted line shows the phoneme border.



Total number of responses
/ae/ = 48%
/a/ = 52%

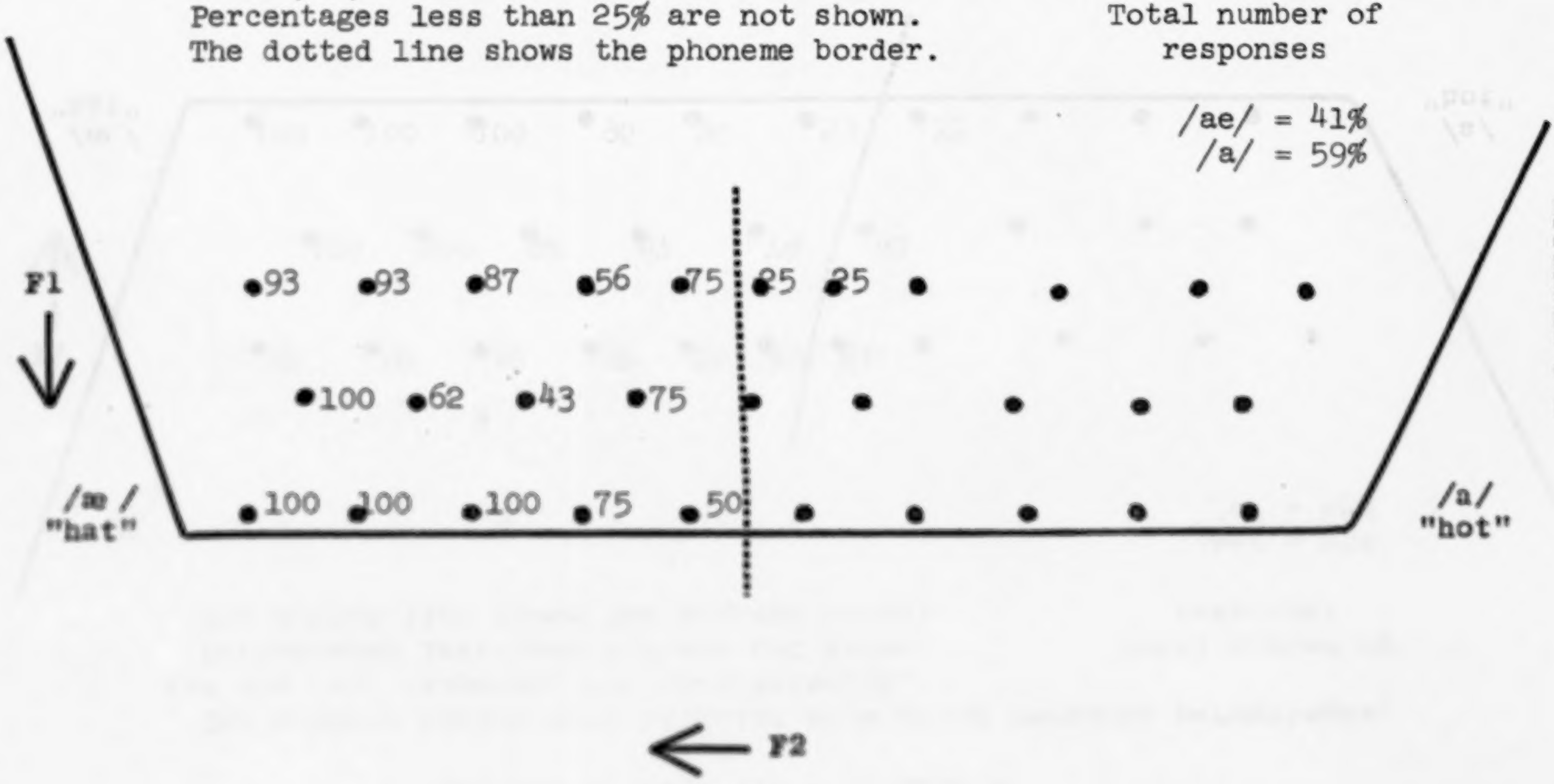
APPENDIX J-2.

Results of Test IIC - Buffalo B

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.



APPENDIX J-3

Results of Test IIC - Ontario A

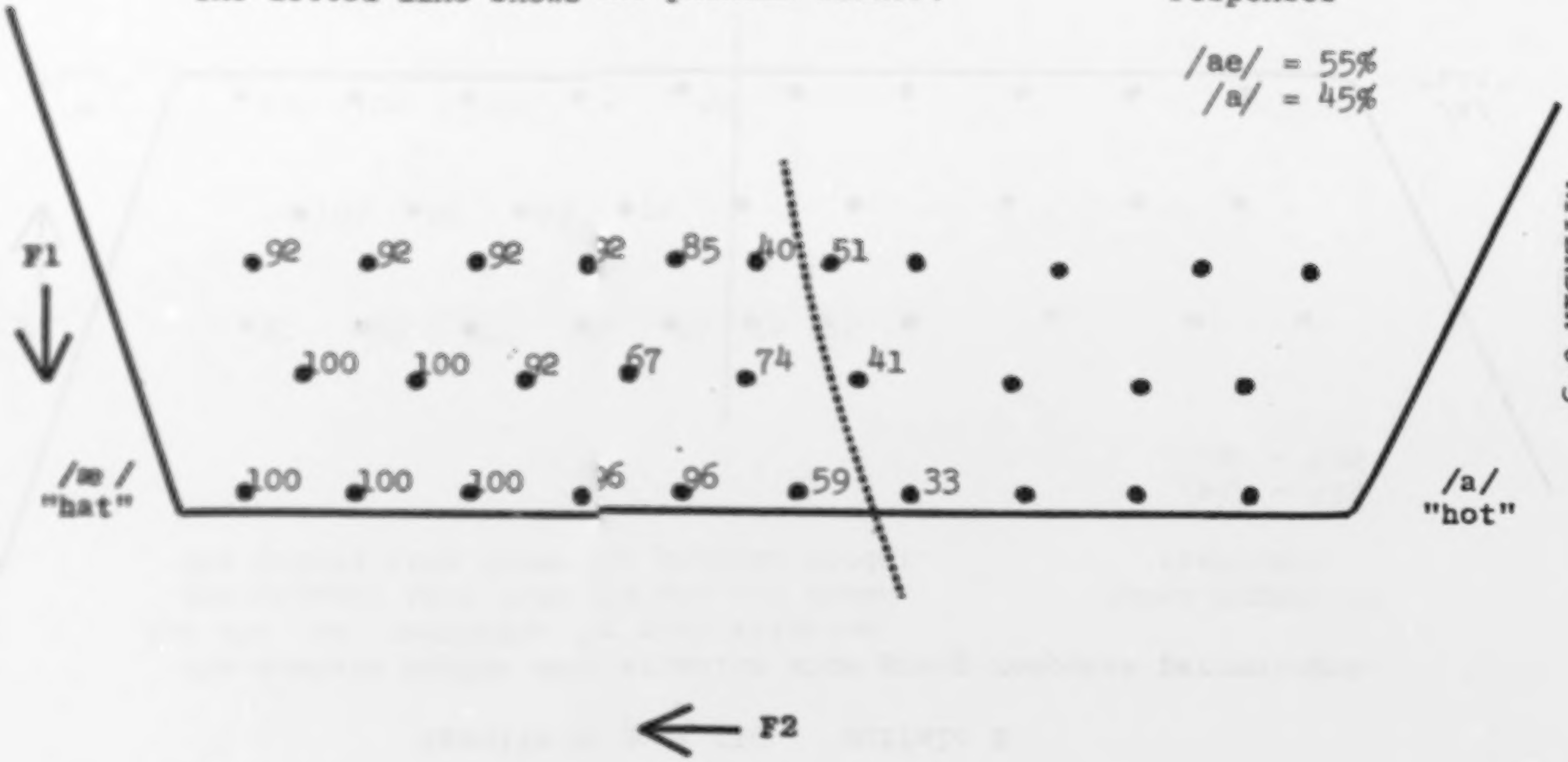
The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus.

Percentages less than 25% are not shown.

The dotted line shows the phoneme border.

Total number of responses

/ae/ = 55%
/a/ = 45%



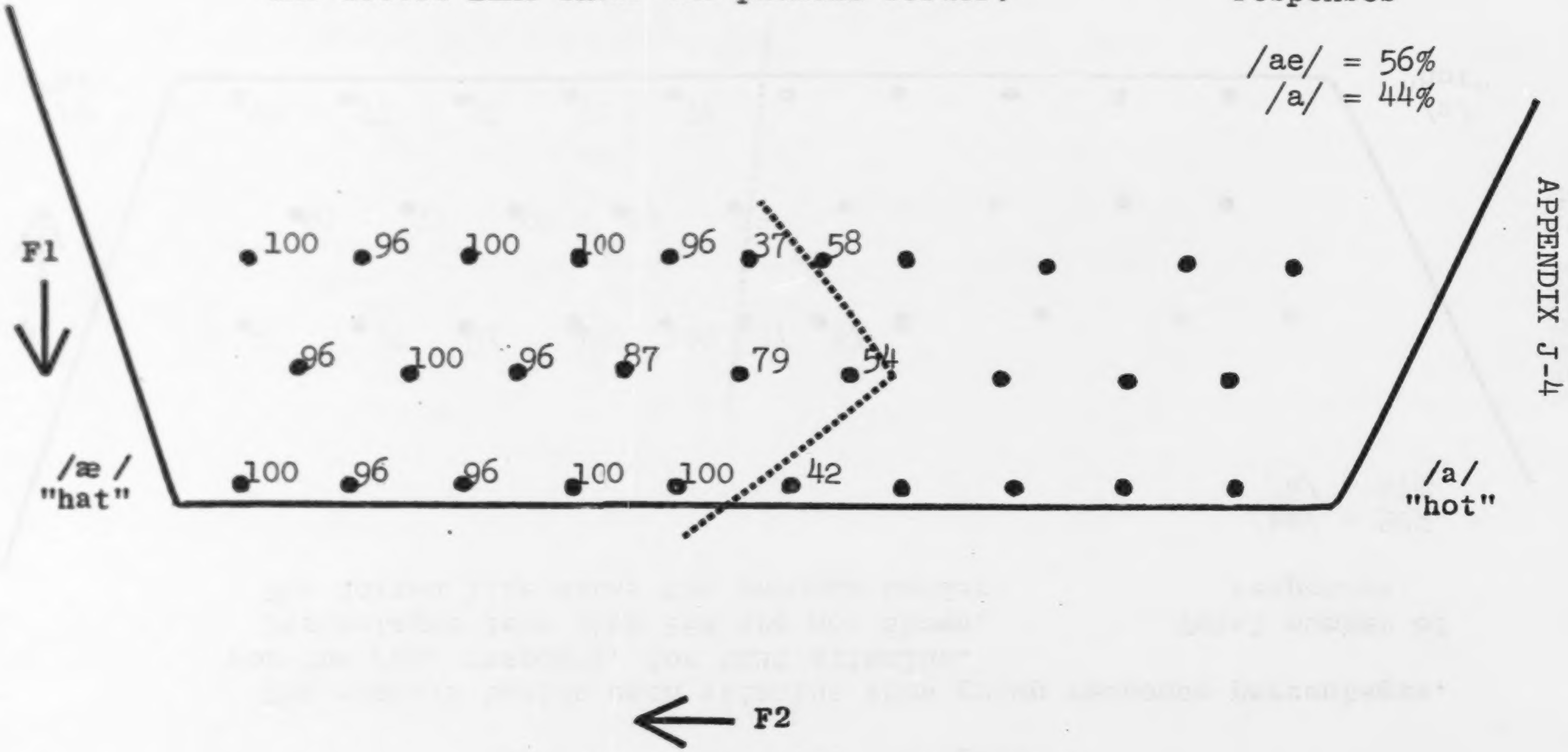
APPENDIX J-4

Results of Test IIC - Ontario B

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus.
Percentages less than 25% are not shown.
The dotted line shows the phoneme border.

Total number of responses

/ae/ = 56%
/a/ = 44%



APPENDIX J-5

Results of Test IIC - Virginia A

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus.

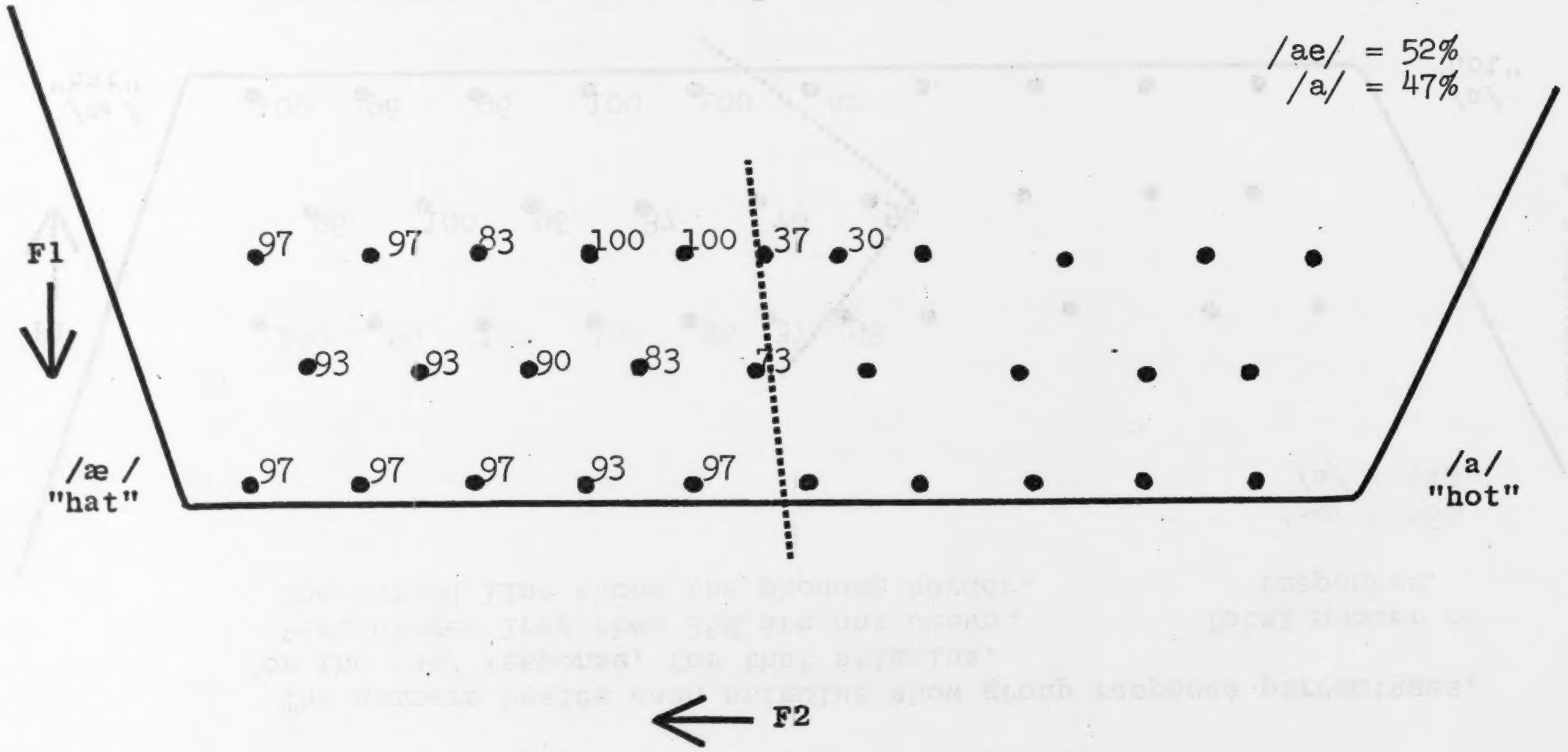
Percentages less than 25% are not shown.

The dotted line shows the phoneme border.

Total number of responses

/ae/ = 52%

/a/ = 47%



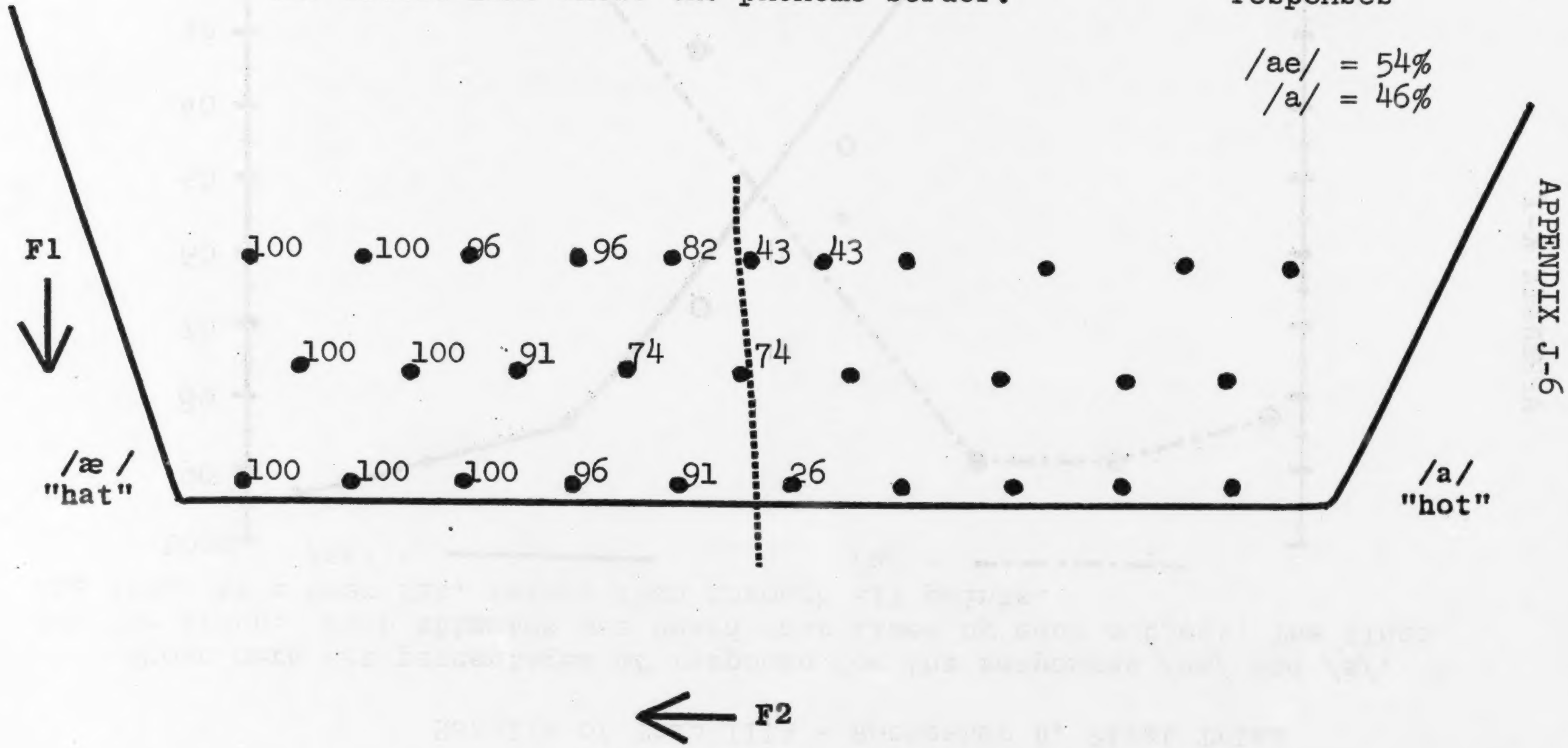
APPENDIX J-6

Results of Test IIC - Virginia B

The numbers beside each stimulus show group response percentages, for the /ae/ response, for that stimulus.
Percentages less than 25% are not shown.
The dotted line shows the phoneme border.

Total number of responses

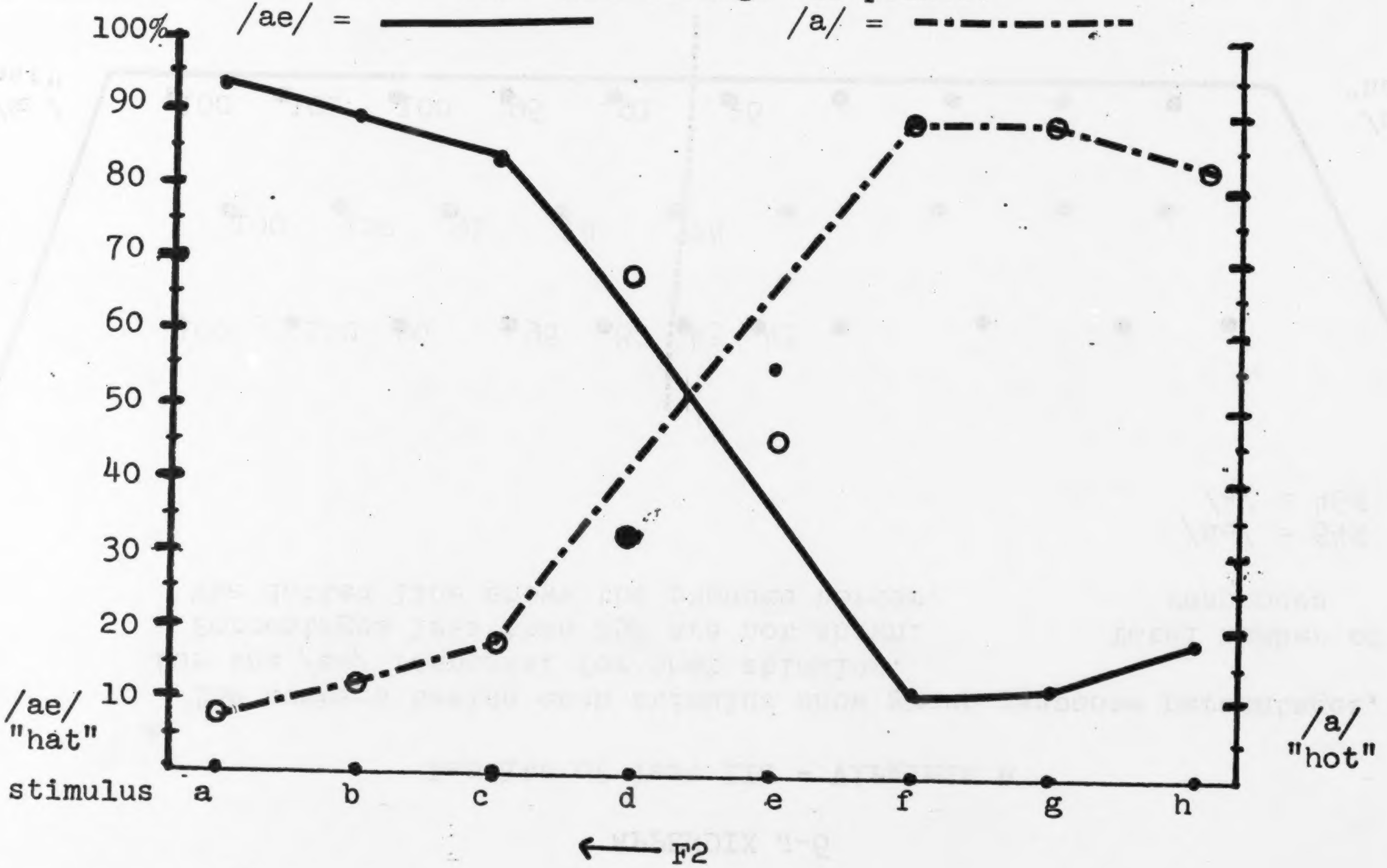
/ae/ = 54%
/a/ = 46%



APPENDIX K-1

Results of Test IIIA - Rochester A, First Trial

Shown here are percentages of response for the responses /ae/ and /a/, for the group. Each stimulus was heard four times by each subject. The lines are drawn as a best fit, rather than through all points.

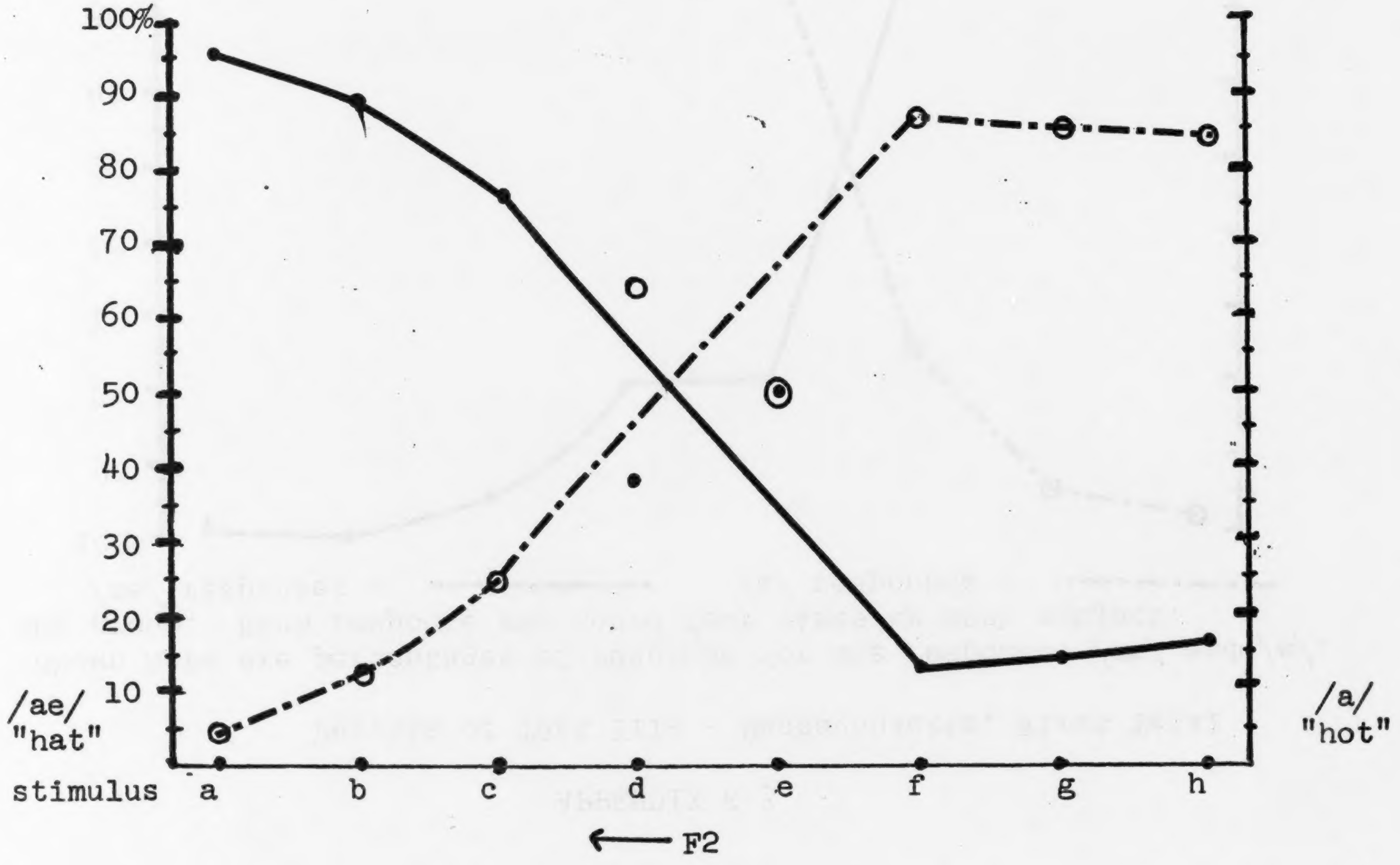


APPENDIX K-2

Results of Test IIIA - Rochester A, Second Trial

Shown here are percentages of response for the responses /ae/ and /a/, for the group. Each stimulus was heard four times by each subject. The lines are drawn as a best fit, rather than through all points.

/ae/ = ———— /a/ = -·-·-·-

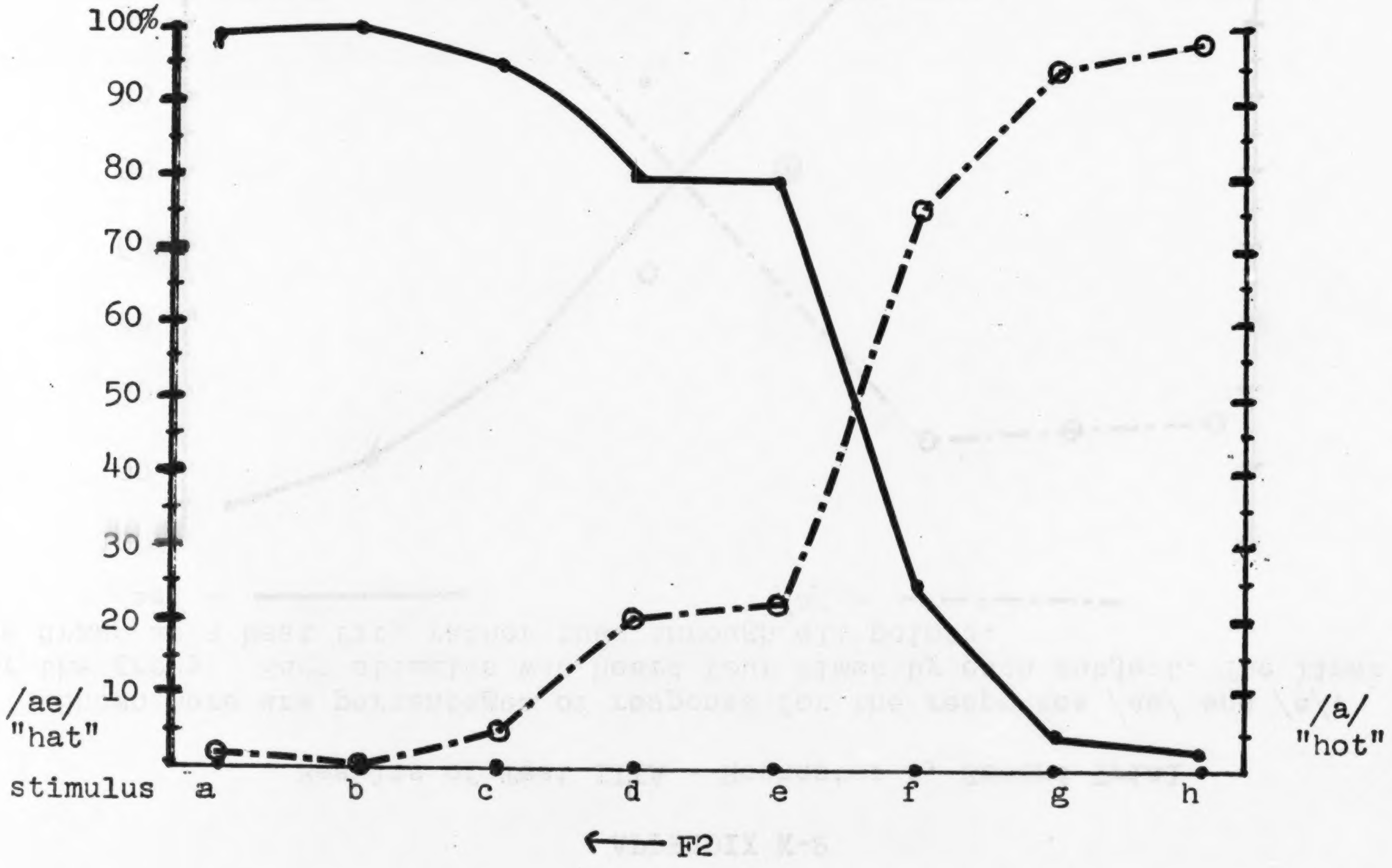


APPENDIX K-3

Results of Test IIIA - Massachusetts, First Trial

Shown here are percentages of response for the responses /ae/ and /a/, for the group. Each response was heard four times by each subject.

/ae/ responses = ———— /a/ responses = -·-·-·-



APPENDIX K-4

Results of Test IIIA - Massachusetts, Second Trial

Shown here are percentages of response for the responses /ae/ and /a/, for the group. Each response was heard four times by each subject.

/ae/ responses = ———— /a/ responses = —·—·—·—·